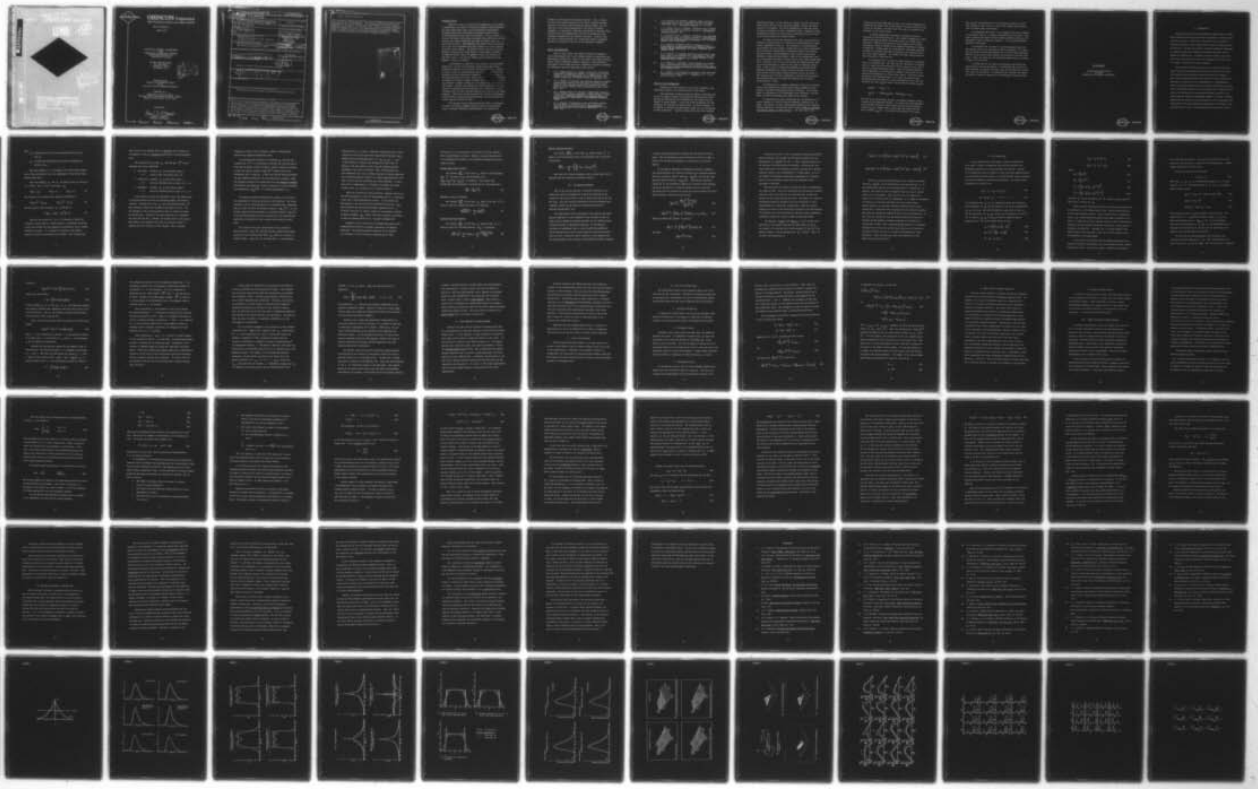
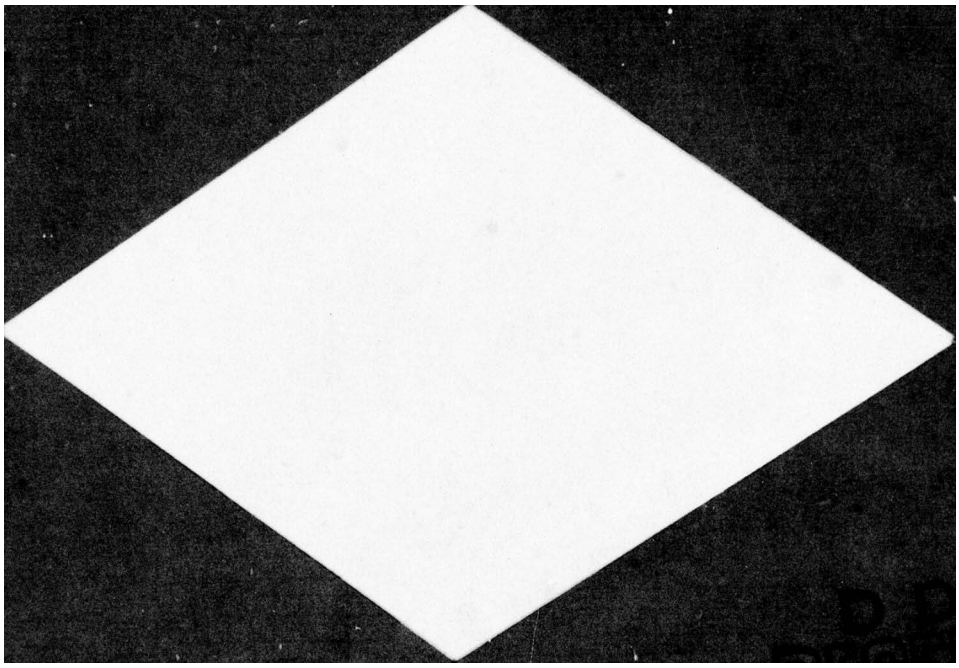
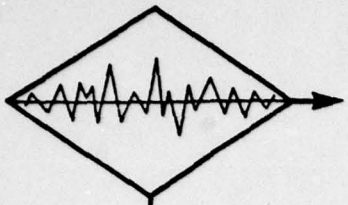
AD-A069 980 ORINCON CORP LA JOLLA CALIF F/G 12/1
PRACTICAL CONTROL ALGORITHMS FOR NONLINEAR STOCHASTIC SYSTEMS A--ETC(U)
MAR 79 D L ALSPACH F44620-75-C-0023
UNCLASSIFIED OC-R-79-0002AA-1 AFOSR-TR-79-0647 NL

| OF |
AD
A069980



END
DATE
FILMED
-7--79
DDC





ORINCON Corporation

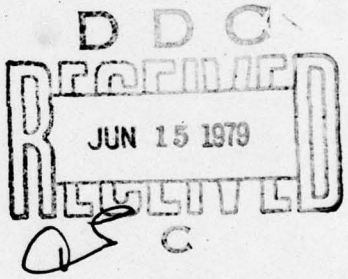
3366 No. Torrey Pines Ct., Suite 320, La Jolla, CA 92037 (714) 455-5530

OC-R-79-0002AA-1

March 1979

PRACTICAL CONTROL ALGORITHMS FOR NONLINEAR STOCHASTIC SYSTEMS AND INVESTIGATIONS OF NONLINEAR FILTERS

Annual Technical Report
For The Period
1 February 1978 to
31 January 1979



Prepared Under
Contract No. F44620-75-C-0023
for the
Air Force Office of Scientific Research

Submitted to:

Department of the Air Force
Air Force Office of Scientific Research (AFSC)
Bolling Air Force Base, D. C. 20332

Prepared By

Daniel L. Alspach
Principal Investigator

Operations Research INformation CONtrol

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: AFOSR-TR- 79-0647	2. GOVT ACCESSION NO. / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): PRACTICAL CONTROL ALGORITHMS FOR NONLINEAR STOCHASTIC SYSTEMS AND INVESTIGATIONS OF NONLINEAR FILTERS.	5. TYPE OF REPORT & PERIOD COVERED: Interim 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Daniel L. Alspach	8. CONTRACT OR GRANT NUMBER(s): F44620-75-C-0023
9. PERFORMING ORGANIZATION NAME AND ADDRESS: ORINCON Corporation, 3366 North Torrey Pines Court, Suite 310, La Jolla, CA 92037	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F / 2304 / A1
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research/NM, Bolling AFB, Washington, DC 20332	12. REPORT DATE: March 1979 13. NUMBER OF PAGES: 74
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): OC-R-79-0002AA-1	15. SECURITY CLASS. (of this report): UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited.	76p.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report): Annual technical rept. 1 Feb 78-31 Jan 79,	
18. SUPPLEMENTARY NOTES	
19. KEY WORDS (Continue on reverse side if necessary and identify by block number): Non-Linear Estimation Stochastic Control; Extended Kalman Filter	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This Annual Technical Report summarizes a continuation of the investigation into the use of digital nonlinear filters in conjunction with deterministic control algorithms. The problem of stabilization and control of nonlinear stochastic systems observed by noisy measurement data arises in many Air Force systems. Inherent in this problem is the problem of processing noise-contaminated measurement data to obtain accurate estimates of the state of the system. If it is possible to estimate the state of the system accurately,

DD FORM 1 JAN 73 1473 UNCLASSIFIED

392 776

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract continued.

then well-known classical deterministic control techniques may often be used to give adequate system performance. This approach will greatly reduce the complexity of the control algorithm over that required by a truly "optimal" stochastic control policy. On the other hand, the use of recently developed filtering techniques in place of the simpler linearized or extended Kalman filter can greatly increase the accuracy of the state estimates and, thereby, improve system performance and alleviate divergence problems.

INTRODUCTION

Inherent in many Air Force system appliations is the problem of stabilization and control of nonlinear stochastic systems observed by noisy measurement data, and the difficulties encountered in processing this noise-contaminated measurement data to obtain accurate estimates of the state of the system. If it is possible to estimate the state of the system accurately, then well-known classical deterministic control techniques may often be used to give adequate system performance. This approach will greatly reduce the complexity of the control algorithm over that required by a truly "optimal" stochastic control policy. On the other hand, the use of recently developed filtering techniques in place of the simpler, linearized or extended Kalman filters can greatly increase the accuracy of the state estimates and thereby improve system performance and alleviate divergence problems.

A straightforward approach to the stochastic control problem would be to use the recent research results giving approximate a posteriori densities in conjunction with the principle of optimality in order to solve the optimal control problem using the well-known stochastic dynamic programming equations. Preliminary studies of this approach show two things. First, even for very simple systems in which it is possible to obtain tractable approximate densities, it is very difficult to obtain solutions for even approximate controls in this manner. Second, even though the control algorithms developed cannot be considered practical because of the extensive calculations involved, they can lead to greatly reduced cost even for very simple problems. This second factor does make the continued investigation of the stochastic control problem interesting, but the first indicates another approach should be used.

The approach taken in this contract has been to investigate the effect of already developed nonlinear filters in the feedback loops of nonlinear stochastic control systems to be followed by

optimal or nearly optimal deterministic controls. This, in effect, can force the separate structure of the linear quadratic Gaussian problem on the more general stochastic control problem, but will make use of the latest developments in the field on nonlinear estimation. In particular, nonlinear measurements of basically linear systems will be investigated since the cost of transducers often increases markedly from the linearity constraint. It may be that the cost of a small computer in a control loop could be completely saved by a reduction in the cost of the transducers required.

WORK ACCOMPLISHED

Work on this contract has allowed continued investigation into the development and use of nonlinear filters in conjunction with deterministic control laws. Much of the detail on the work performed under Contract F44620-75-C-0023 is contained in the publications listed below, all of which have been published, accepted or submitted for publication since the beginning of this contract.

1. D. L. Alspach and R. N. Lobbia, "A score for correct data association in multitarget tracking," to appear in an invited session in the 1979 Proceedings of the IEEE Decision and Control Conference in Orlando, Florida, in December, 1979.

2. D. L. Alspach, "A Gaussian sum Bayesian approach to passive bearings-only tracking," invited paper, Proceedings of the Office of Naval Research Sponsored Conference on Target Motion Analysis, Monterey Naval Postgraduate School, 25-27 May, 1977.

3. D. L. Alspach and H. W. Sorenson, "Approximate solutions of the nonlinear filtering equations," invited chapter in forthcoming book, Nonlinear Estimation and Filtering Theory: A Status Review, E. Stear (Ed.), to be published by Marcel Dekker.

4. D. L. Alspach, "A discussion of the relationships between the dual goals of stochastic control," An International Journal of Computers and Electrical Engineering, 4:1, January, 1977.

5. D. L. Alspach, "A stochastic regulator using a certainty equivalence control with a nonlinear filter for processing hard-limited data," Information Sciences, Vol. 13, 1978.

6. L. L. Scharf and D. L. Alspach, "Nonlinear state estimation in observation noise of unknown covariance," International Journal of Control, 1978.

7. L. L. Scharf and D. L. Alspach, "Nonlinear state estimation in observation noise of unknown covariance," Proceedings of the 1976 Joint Automatic Control Conference (West Lafayette, Indiana, July 27-30, 1976).

8. D. L. Alspach, "Nonlinear filters in feedback control," Proceedings of the Sixth Symposium on Nonlinear Estimation Theory and Its Applications, San Diego, California (September, 1975).

9. D. L. Alspach, "A certainty equivalence control with a nonlinear filter in the feedback loop," Proceedings of the 1975 IEEE Symposium on Decision and Control, Houston, Texas (December 10-12, 1975).

10. D. L. Alspach, "A stochastic control algorithm for systems with control dependent plant and measurement noise," An International Journal of Computers and Electrical Engineering, 2:4, November, 1975.

11. D. L. Alspach, "A Gaussian sum approach to the multitarget identification-tracking problem," Automatica, Vol. 11, pp. 285-296 (August, 1975).

PUBLICATIONS SUMMARIES

Following are brief summaries of the work contained in the publications listed in the previous section of this report.

The multitarget tracking problem is defined in Publication #1 as that of taking a number of measurements obtained from several sensors and determining track estimates for any targets that are "heard" by these sensors. In the real world environment the measurements are cluttered by random noise. In these situations, it is difficult to determine precisely which target (if any) corresponds to each measurement. Typical problems which arise with tracking

algorithms include: too few tracks are formed; too many tracks are formed (false tracks); and, inaccurate position, course, and speed estimates are reported. The above difficulties are often the result of incorrect allocation of data to individual tracks. Algorithms, while estimating the motion of a given target, inadvertently mix in clutter and/or measurements from another target.

In order for a correct allocation to be made, we must have an effective scoring formula; i.e., some means of determining how likely a given assignment of data is. To be effective, a scoring formula must produce (on the average) a better score for correct assignments than for incorrect assignments. Information useful in the scoring process includes a priori intelligence data (such as initial target locations), models of target motion, models of the transmission channel, and expected moments of clutter for the sensor gain setting being used. Basically, the score is derived from the residuals which come out of the processing of a batch of data with the extended Kalman filter. This is used to evaluate the likelihood of potential tracks. Although "likelihood" has a useful intuitive meaning, we use the term to mean the probability density function $p(\lambda)$ of the track λ. The expected cost of a given assignment is derived with the theory of extremals being used to obtain the expected cost of adding a clutter point in a track.

In Publication #2, a specific application of the use of Gaussian sums to the bearings-only target motion analysis problem was presented at the Naval Postgraduate School in Monterey. This invited paper was published in the Proceedings of the conference, the main theme of which was bearings-only target motion analysis.

A summary/review paper (Publication #3) was prepared as an invited chapter of a text on nonlinear estimation, edited by Dr. E. Stear. The paper is entitled "Approximate Solutions of the Nonlinear Filtering Equations" and the book will be entitled Nonlinear Estimation and Filtering Theory: A Status Review. The work done on the

contract to date including that in many of the above publications is summarized in detail in this review chapter, as well as work by other workers in the field. For this reason this paper is attached as an appendix to this annual report.

A general philosophical approach to stochastic control is discussed in Publication #4 above. The method of aligning the "dual" goals of the general optimal stochastic control as a design tool is discussed. When the two goals are exactly aligned, the certainty equivalence control is optimal and no additional intentional "probing" is required. If these goals are "anti-aligned," demanding opposing controls, the certainty equivalence control can be, locally at least, the worst control possible. An example with this characteristic has been discussed in Publication #4.

In Publication #5 we consider a simple example of a nonlinear filter in a feedback loop. For this case the "dual goals" are aligned and it is shown that the performance of the system is very close to that of an optimal stochastic control. Since the optimal stochastic control algorithm is very difficult to calculate, this is done by comparing the performance to a known lower bound. This lower bound is found in the following manner. It is clear that there is more information about the state in a linear measurement of the state contaminated by noise $z_{LIN}(k)$ than in the hard-limited version $y_{HL}(k)$:

$$z_{LIN}(k) = H_k x_k + v_k$$

$$y_{HL}(k) = \text{SIGN}(z_{LIN}(k)) = \text{SIGN}(H_k x_k + v_k).$$

With the linear measurement function, the optimal stochastic control is given by the separation theorem. It is clear that the performance of the optimal stochastic control system with only the hard-limited function $y_{HL}(k)$ available will be worse than the system with the linear function $z_{LIN}(k)$. It is shown in Publication #4 that given

only $y_{HL}(k)$, the performance of the certainty equivalence control with a nonlinear filter is close to that of the optimal control system with the better linear measurement.

In Publications #6 and #7, a new nonlinear filter was developed in conjunction with Dr. L. Scharf of Colorado State University. This paper considers the adaptive Kalman filtering problem where only the measurement noise covariance is unknown. A new parallel filtering algorithm is developed.

In Publication #8, the effect of control-dependent plant and measurement noise on the feedback control is discussed. It is shown that the goals are effectively aligned, but not of the same magnitude. The use of any control action reduces the accuracy of the state estimates. Thus, the effect of such control-dependent noise is to induce "caution" on the control.

In Publication #9, the filter for processing hard-limited measurement data was first introduced. In Publication #10, a filter for state estimation in systems with state- and control-dependent measurement noise was introduced. In Publication #11, Gaussian sum filters are used in a Bayesian approach to the multitarget tracking problem.

ATTACHMENT

APPROXIMATE SOLUTIONS OF THE NONLINEAR FILTERING EQUATIONS

I. INTRODUCTION

Many physical systems are described most appropriately by mathematical models which take into account the stochastic influences which act upon the system. Furthermore, the behavior of the system can seldom be observed in such a manner that its state is known precisely at each time. Instead, the system is observed through the measurement of variables which yield incomplete information and which contain unknown or random errors. In general one must estimate, from these noisy data, the state of the system and, frequently, take some type of "control" action based on these estimates.

The problem of estimating the state of a nonlinear stochastic system from noisy measurement data has been the subject of considerable research interest during the past few years but, although a great deal has been published on the subject, the basic objective of obtaining a solution that can be implemented in a straightforward manner for specific applications has not been satisfactorily realized. This is manifested by the fact that the Kalman filter equations [1,2], which are valid only for linear, gaussian systems, continue to be widely used for nonlinear, nongaussian systems. Of course, continued application has resulted in the development of ad hoc techniques, discussed below, that have improved the performance of the Kalman filter and which give it some of the characteristics of nonlinear filters.

Central to the nonlinear estimation and stochastic control problems is the determination of the probability density function of the state conditioned on the available measurement data. If this a posteriori density function were known, an estimate of the state for any performance criterion could be determined. Unfortunately, although the manner in which the density evolves with time and additional measurement data can be described in terms of difference (or differential) equations, these relations are generally very difficult to solve, either in closed-form or numerically, so that it is usually impossible to determine the a posteriori density for specific applications. Because of this difficulty, it is natural to investigate the possibility of approximating the density with some tractable form.

II. THE GENERAL PROBLEM

Many formulations of stochastic control problems are possible and have appeared in the literature. If the objective is to determine computational algorithms, it is reasonable that attention should be directed toward the development of control and estimation policies that explicitly assume that events occur at discrete instants of time.

Suppose that the state vector $\underline{x}$ of the system evolves according to the nonlinear stochastic difference equation:

$$\underline{x}_{k+1} = \underline{f}_k \left(\underline{x}_k, \underline{u}_k, \underline{w}_k \right) \quad ; \qquad k = 0, 1, \ldots, N-1 \tag{1}$$

where

$\underline{x}_k \sim$ n-dimensional state of the system at the time t_k;

$\underline{u}_k \sim$ p-dimensional control vector that acts on the system for $t_k \leq t < t_{k+1}$;

$\underline{w}_k \sim$ q-dimensional plant random noise vector that acts on the system for $t_k \leq t < t_{k+1}$;

The random noise sequence* $(\underline{w}_0, \underline{w}_1, \ldots, \underline{w}_k) \triangleq \underline{w}^k$ is assumed to have a known probability distribution $p(\underline{w}^k)$ such that the $\underline{w}_i$ are independent between sampling times [i.e., $p(\underline{w}_o, \underline{w}_1, \ldots, \underline{w}_k) = p(\underline{w}_o), p(\underline{w}_1), \ldots, p(\underline{w}_k)$ for all k]. Sequences having this characteristic will be referred to as white-noise sequences. The initial state $\underline{x}_0$ is also considered to be random variable with a known distribution and is taken to be independent of the plant noise.

The behavior of the plant (1) is observed imperfectly through measurement quantities $\underline{z}_k$ that are related in a prescribed fashion to the state but which contain random errors.

$$\underline{z}_k = \underline{h}_k(\underline{x}_k, \underline{v}_k) \; ; \qquad k = 0, 1, \ldots, N. \tag{2}$$

---

* Throughout this work a vector or scalar with an algebraic superscript (k) will mean the total array of such vectors which have occurred at all times up to and including t_k.

where

$\underline{z}_k \sim$ m-dimensional vector of known measurement data at the time t_k;

$\underline{v}_k \sim$ r-dimensional measurement noise vector contaminating the data at t_k.

The noise sequence $\underline{v}^k$ is assumed to be a white-noise sequence with a known distribution and to be independent of the initial state and all plant noise.

The noise sequence $\underline{v}_k$ and $\underline{w}_k$ are taken to have zero mean and the initial state vector to have mean $\hat{\underline{x}}_o'$.

$$E(\underline{x}_o) = \hat{\underline{x}}_o' \quad , \qquad E(\underline{v}_k) = \underline{0} \quad , \qquad E(\underline{w}_k) = \underline{0} \tag{3}$$

The covariance of the white-noise sequence and defined by

$$E(\underline{w}_k \, \underline{w}_j^T) = Q_k \, \delta_{kj} \quad , \qquad E(\underline{v}_k \, \underline{v}_j^T) = R_k \, \delta_{kj} \tag{4}$$

and the initial state covariance P_o' is defined by:

$$E\left((\underline{x}_o - \hat{\underline{x}}_o')(\underline{x}_o - \hat{\underline{x}}_o')^T\right) = P_o' \tag{5}$$

Based on the system (1) - (2), it is possible to define the stochastic control problem. Before doing so, it should be noted that it has been assumed that the probability distributions for all random variables are known. It is possible to consider a more general problem in which the distributions are unknown. This situation has

been referred to by Bellman [3] as an adaptive control problem to distinguish it from the stochastic problem that is being formulated here.

The estimation of the state $\underline{x}_k$ from the data $\underline{z}^{k+\gamma}$ can be separated into three subproblems:

1. Filtering: estimate $\underline{x}_k$ at the present stage k
   $\gamma = 0$ using all past and current data $\underline{z}^k$.
2. Prediction: estimate $\underline{x}_k$ at some future stage k
   $\gamma < 0$ using all available data at stage $k+\gamma$, $\gamma < 0$.
3. Smoothing: estimate $\underline{x}_k$ at some earlier stage k
   $\gamma > 0$ using all available data at stage $k+\gamma$, $\gamma > 0$.

In the absence of plant and measurement noise, the problem that is considered below would have the following simple, deterministic statement. Determine the state $\underline{x}_k$ from the measurement data $\underline{z}^k$. When stochastic effects are included in the model as in (1) - (2), the statement of the filtering problem comes to include an element of arbitrariness. Certainly, the basic objective is to "estimate" the state from the data. However, the random effects in the system model implies that redundant data must be collected in order to minimize the noise influence on the estimate. Now, it becomes

reasonable to define "best" estimates, thereby introducing the arbitrariness mentioned immediately above.

In considering the question of estimating $\underline{x}_k$ from the measurement data $\underline{z}^{k+\gamma}$, it is necessary to specify the criterion that is to be used to define a "best" estimate. However, whatever criterion is used, the density function $p(\underline{x}_k / \underline{z}^{k+\gamma})$ contains all of the information that is required. In fact this density function provides the most complete description of the system that is possible. Thus, the estimation problem can be approached from this Bayesian viewpoint [4,5] without specifying the criterion because one is first concerned with determining $p(\underline{x}_k / \underline{z}^{k+\gamma})$ or a valid approximation to it.

The Bayesian approach described below applies to all three of the subproblems of filtering, prediction, and smoothing. The actual calculation of the smoothing density is considerably more complicated than the other two while the prediction density $p(\underline{x}_k / \underline{z}^{k+\gamma})$ $(\gamma < 0)$ follows simply from the filtering of a posteriori density $p(\underline{x}_k / \underline{z}^k)$. In the following discussion we will specialize to the filtering and prediction cases and most of the research results apply to these cases.

For systems of the type considered here the a posteriori density function $p(\underline{x}_k / \underline{z}^k)$ provides the most complete description possible of the so-called state vector $\underline{x}_k$ which is, of course, a random variable. $p(\underline{x}_k / \underline{z}^k)$, on the other hand, is a deterministic

function which is, in theory, completely determined by the a priori statistics of the noise and initial state density function $p(\underline{x}_o)$, combined with the measurement data $\underline{z}^k = (\underline{z}_1, \underline{z}_2, \underline{z}_3, \ldots, \underline{z}_k)$. If this function is taken as the description of the state, it reduces to the unit impulse at the true state whenever perfect knowledge of the state is obtained. Thus, accepting this as a valid definition of state, it becomes necessary to obtain explicitly the a posteriori density function or a "good" approximation to it in order to solve both the estimation and control problems. In fact, once this is accomplished it is possible to estimate the random variable state $\underline{x}_k$ according to any criterion function.

While the a posteriori density provides a complete solution of the filtering problem, it has the disadvantage that it is a function rather than a finite-dimensional estimate. If the problem were deterministic, the solution would be provided by the vector $\hat{\underline{x}}_{k|k}$ that satisfied the plant and measurement equations for all k. A similar "solution" is commonly sought for the stochastic problem. To obtain estimates $\hat{\underline{x}}_{k|k}$, useful, but often arbitrary, performance criteria are defined which lead to "optimal" estimates [6].

Examples of such criteria are the minimum mean square error, minimum absolute deviation, maximum a posteriori, and maximum likelihood. The Bayesian approach yields all of the information necessary to obtain estimates satisfying any of these

criteria so it is not necessary at this point to be more specific about the performance criterion. However, it may be instructive to see a mathematical statement of the estimation problem for the more common criteria.

MINIMUM MEAN-SQUARE ESTIMATE

The estimate $\hat{\underline{x}}^{MS}_{k|k}$ of the state $\underline{x}_k$ based on the measurement data $\underline{z}^k$ is chosen so that the mean-square error $E\left[(\underline{x}_k - \hat{\underline{x}}_{k|k})^T (\underline{x}_k - \hat{\underline{x}}_{k|k})\right]$ is minimized. The estimate that accomplishes the minimization is provided by the conditional mean

$$\hat{\underline{x}}^{MS}_{k|k} = E\left[\underline{x}_k | \underline{z}^k\right] . \tag{6}$$

MAXIMUM A Posteriori ESTIMATE

The estimate $\hat{\underline{x}}^{MAP}_{k|k}$ of the state $\underline{x}_k$ based on the data $\underline{z}^k$ is chosen so that the a posteriori density is maximized.

$$p\left(\hat{\underline{x}}^{MAP}_{k|k} \middle| \underline{z}^k\right) = \max_{\underline{x}_k} p\left(\underline{x}_k \middle| \underline{z}^k\right) \tag{7}$$

MAXIMUM LIKELIHOOD ESTIMATE

The estimate $\hat{\underline{x}}^{ML}_{k|k}$ of the state $\underline{x}_k$ based on data $\underline{z}^k$ is chosen so that the likelihood function $\lambda(\underline{x}_k)$ is maximized.

$$\lambda\left(\hat{\underline{x}}^{ML}_k, \underline{z}_k\right) = \max_{\underline{x}_k} p(\underline{z}_k | \underline{x}_k) = \max_{\underline{x}_k} \frac{p(\underline{x}_k | \underline{z}_k)\ p(\underline{z}_k)}{p(\underline{x}_k)} \tag{8}$$

MINIMUM ABSOLUTE DEVIATION

The estimate $\hat{\underline{x}}^{MAD}_{k \mid k}$ of the state $\underline{x}_k$ based on data $\underline{z}^k$ is chosen so that the expected value of the absolute value of the error is minimized.

$$\hat{\underline{x}}^{MAD}_{k \mid k} = \min_{\hat{\underline{x}}_k} E\left[\sum_{i=1}^{n} |x_{ik} - \hat{x}_{ik|k}| \Big/ \underline{z}^k\right]$$

Note that all of these estimates as well as many others can be obtained from the a posteriori deviation function.

III. THE BAYESIAN APPROACH

Much of the current research on nonlinear filtering is concerned with recursive formulations in which the solution for the *solution of the (k-1)th stage is used to obtain the solution for the* kth stage. Only the recursive formulation shall be considered here. A general solution of the recursive filtering problem can be obtained through Bayes' rule.

The major feature which distinguishes this approach from other possible approaches is the assumption of the existence of well defined a priori probability density functions for all unknown vectors entering the plant or measurement equations. In the Bayesian procedure the measurement data is used to modify the probability density function of the state vector based on all previous measurements and this a posteriori density function is used together with the known dynamical plant and plant noise probability density function

to obtain the predicted density function for the state at the next stage. Thus the probability density function for state at stage k based on all available measurements is calculated in a recursive fashion.

In the Bayesian approach to determining recursive estimation and control policies for stochastic systems one is concerned primarily with the a posteriori density function, $p(\underline{x}_k|\underline{z}^k)$, and the one stage predicted density function, $p(\underline{x}_{k+1}|\underline{z}^k)$. These density functions contain all of the information required for solution of both problems, and can be described by a recursion relation that is useful for obtaining recursive filters and closed loop control policies [4,5,7,8]. These recursion relations are given below:

$$p\left(\underline{x}_k|\underline{z}^k\right) = \frac{p\left(\underline{x}_k|\underline{z}^{k-1}\right) p(\underline{z}_k|\underline{x}_k)}{p\left(\underline{z}_k|\underline{z}^{k-1}\right)} \tag{10}$$

$$p\left(\underline{x}_k|\underline{z}^{k-1}\right) = \int p\left(\underline{x}_{k-1}|\underline{z}^{k-1}\right) p\left(\underline{x}_k|\underline{x}_{k-1}, \underline{u}_{k-1}\right) d\underline{x}_{k-1} \tag{11}$$

where the normalizing constant is given by:

$$p\left(\underline{z}_k|\underline{z}^{k-1}\right) = \int p\left(\underline{x}_k|\underline{z}^{k-1}\right) p(\underline{z}_k|\underline{x}_k)\, d\underline{x}_k \tag{12}$$

and where

$$p\left(\underline{x}_0|\underline{z}^{-1}\right) \triangleq p(\underline{x}_0) \tag{13}$$

Utilization of (10) to (12) in conjunction with the prescribed initial condition (13) provides the information required for the determination of $p(\underline{x}_k|\underline{z}^k)$ for any k. Thus, a general solution of the nonlinear filtering problem is available. Unfortunately, the actual evaluation of the Bayesian recursion relations for a specific nonlinear system is not accomplished in a trivial manner. It is to the problem of developing computational algorithms for evaluating (10) to (13) for specific systems that the remainder of this discussion is directed.

When the system is nonlinear or when the noise is nongaussian, two problems arise. First the integrations in the Bayesian recursion relations cannot be carried out in closed-form and, second, the moments are not easily obtained from (10). The moments are useful in establishing practical estimation and control policies so their determination is an important consideration. These two aspects pinpoint the source of the difficulties involved in the determination of estimation and control policies for nonlinear and/or nongaussian stochastic systems when trying to apply the Bayesian method.

The densities $p(\underline{z}_k|\underline{x}_k)$ and $p(\underline{x}_k|\underline{x}_{k-1})$ can be written more explicitly in the special case that the noise terms $\underline{w}_k$ and $\underline{v}_k$ are assumed to be gaussian and to enter equation (1) and (2) in an additive fashion. Then by assuming that Q_k^{-1} and R_k^{-1} exist, we can write these densities as

$$p(\underline{z}_k|\underline{x}_k) = k_v \exp\left\{-\tfrac{1}{2}[\underline{z}_k - \underline{h}_k(\underline{x}_k)]^T R_k^{-1}[\underline{z}_k - \underline{h}_k(\underline{x}_k)]\right\} \quad (14)$$

$$p(\underline{x}_{k+1}|\underline{x}_k) = k_w \exp\left\{-\tfrac{1}{2}[\underline{x}_{k+1} - \underline{f}_k(\underline{x}_k)]^T Q_k^{-1}[\underline{x}_{k+1} - \underline{f}_k(\underline{x}_k)]\right\} \quad (15)$$

Given the a priori density functions, the a posteriori density functions $p(\underline{x}_k|\underline{z}^k)$ can be determined for any sampling time t_k if the integrations required can be accomplished in a closed-form. In general, this cannot be done and suggests that some type of approximation must be considered. When only the first two moments are known, or the initial density is nongaussian, it is common to approximate the density as a gaussian with these first and second moments. Another method is to linearize nonlinear problems around a known nominal and to assume the noise is gaussian in the linearized problem. The reason for wanting the problem in this form is well known [2]. In this case the a posteriori and the predicted density functions are Gaussian and the Bayesian recursion relations can be solved in a closed form. In fact, since a Gaussian density function is completely determined by its first and second order statistics, the functional recursion relation reduces to a recursion relationship for these statistics. These relations have come to be referred to as the Kalman filter equations [1,2].

IV. THE LINEAR CASE

In many applications of the theory a nominal trajectory in state space is known or assumed and the nonlinear dynamical and measurement equations are linearized about this nominal. Because of the comparative ease of solution of this linearized problem relative to the nonlinear one and because of its applicability in many instances, this linear problem has received much attention. The linear version of Equations (1) and (2) are:

$$\underline{x}_{k+1} = F_k \quad \underline{x}_k + \Gamma_k \underline{u}_k + \underline{w}_k \tag{16}$$

$$\underline{z}_k = H_k \underline{x}_k + \underline{v}_k \; ; \quad k = 0, 1, \ldots, N. \tag{17}$$

The assumption that $\underline{w}_k$, $\underline{v}_k$ are independent white noise sequences, both independent of $\underline{x}_0$, will be made here, but is not necessary [9]. F_k, H_k, and $\Gamma_k \underline{u}_k$ are known deterministic quantities at time t_k and the statistics of $\underline{w}_k$, $\underline{v}_k$, and $\underline{x}_0$ are all defined in Section II. If, in addition, $\underline{v}_k$, $\underline{w}_k$, and $\underline{x}_0$ are gaussian random variables, equations (10) - (13) can be solved exactly to give:

$$K_k = P_k' H_k^T \left[H_k P_k' H_k^T + R_k \right]^{-1} \tag{18}$$

$$\hat{\underline{x}}_k = \hat{\underline{x}}_k' + K_k \left[\underline{z}_k - H_k \hat{\underline{x}}_k' \right] \tag{19}$$

$$P_k = P_k' - K_k H_k P_k' \tag{20}$$

$$P'_{k+1} = F_k \, P_k \, F_k^T + Q_k \tag{21}$$

$$\underline{\hat{x}}'_{k+1} = F_k \quad \underline{\hat{x}}_k + \Gamma_k \, \underline{u}_k \tag{22}$$

where

$$\underline{\hat{x}}_k = E\left(\underline{x}_k \,|\, \underline{z}^k\right) \tag{23}$$

$$\underline{\hat{x}}'_k = E\left(\underline{x}_k \,|\, \underline{z}^{k-1}\right) \tag{24}$$

$$P_k = E\left[(\underline{x}_k - \underline{\hat{x}}_k)(\underline{x}_k - \underline{\hat{x}}_k)^T \,\middle|\, \underline{z}^k\right] \tag{25}$$

$$P'_k = E\left[(\underline{x}_k - \underline{\hat{x}}'_k)(\underline{x}_k - \underline{\hat{x}}'_k)^T \,\middle|\, \underline{z}^{k-1}\right] \tag{26}$$

and where $\underline{u}_k$ can be any function of $\underline{z}^k$ and the a priori data and thus can be a function of $\underline{\hat{x}}_k$.

These recursion relations are exact for the Gaussian problem and are referred to as the Kalman filter [1,2]. Several characteristics of this filter should be noted. First, the mean of the a posteriori density $\underline{\hat{x}}_k$ always provides the minimum mean-square estimate for the state. In this case of linear systems, when the a priori densities are Gaussian, the mean provides the optimal estimate for a large class of estimation criteria [6]. Secondly, the P matrix arising in the Kalman filter is the covariance of the error in the estimate, and it is independent of all measurements.

If the noise is non-Gaussian and the minimum mean-square error estimate is desired, the Kalman filter still provides the best "linear" estimate for state. In this case, however, estimators with smaller

error variances are possible. This can be seen from the fact that, in general, the variance is a function of the measurements. A simple example showing this is given below.

Consider the scalar one stage plant:

$$z_0 = x_0 + v_0 \tag{27}$$

where both x_0 and v_0 are uniformly distributed on (-1, 1) with variance $\sigma^2 = 1/3$. The approximation of x_0 and v_0 by gaussian random variables gives:

$$p_A(v_0) = N(v_0, 1/3) = \exp(-1.5\ v_0^2) / \sqrt{2\pi/3} \tag{28}$$

$$p_A(x_0) = N(x_0, 1/3) \tag{29}$$

$$p_A(x_0|z_0) = N(x_0 - z_0 / 2, 1/6) \tag{30}$$

Thus giving the linear or Kalman predicted variance of 1/6. The exact value of σ^2 is plotted in Figure 1 versus z_0 and the Kalman approximation to it, $\sigma^2_{Kalman} = 1/6$, is also indicated. With the true distribution of x_0 and v_0 the measurement must be contained in the interval (-2, 2) while for the same system with gaussian noise any measurement is possible.

For both the true and approximate cases the mean of the a posteriori density function is $\frac{1}{2}\ z_0$. This is because the true mean is linear in z_0 for one stage. This is not true for subsequent

stages and as a result the Kalman filter only gives the best linear projection of the best minimum variance estimate for more than one stage. However, as shown in Figure 1, the major problem with the Kalman filter is that the variance calculated by it is not a very realistic approximation of the true variance. This carries over to the nonlinear case and has been one of the major problems of the linearization procedure [10]. Note that in this simple example the true variance can be from twice the Kalman estimate to zero. If one wants to find an estimate for state, even in the linear non-gaussian case, different from minimum variance, then the Kalman estimate does not even represent a best linear estimate. This simple example has been considered at length in references [11,12] where a larger number of stages and where several approximate density functions were compared with the true one. Other cases of linear systems with nongaussian noise and initial states are discussed in references [11-15]. This special case of linear systems with non-gaussian noise definitely requires nonlinear processing of the data in order to form optimal state estimates.

V. APPROXIMATE SOLUTION OF THE BAYESIAN RECURSION RELATIONS

As discussed in the previous section, the only general system for which closed-form solutions of (10) - (13) can be found is when the plant and measurement equations are linear and the statistics of the noise and initial state are Gaussian. Then, the a posteriori density is Gaussian and the conditional mean and covariance are described by the Kalman filter equations. Unfortunately, it is necessary to seek the solution of the Bayesian recursion relations numerically for nonlinear or non-Gaussian systems.

In essence, we are faced with the problem of evaluating multi-dimensional integrals. Certainly, the determination of $p(\underline{x}_k|\underline{z}^{k-1})$ using (11) requires an integration. The calculation of the filtering density $p(\underline{x}_k|\underline{z}^{k-1})$ using (10) is seen to require the multiplication of two density functions. This does not represent a difficult task other than in the storage requirements that are implied in such an operation. However, the normalization and the determination of moments requires integration of the product.

We shall first consider the solution of the problem from a very basic point of view. Clearly, the a posteriori density is a random function of the data. When a measurement realization is available, then we have the density as a function of the state $\underline{x}_k$. To emphasize this and to reduce the notational complexity, let us make the following conventions. The prediction density shall be

written as

$$p(\underline{x}_k | \underline{z}^{k-1}) \triangleq \pi(\underline{x}) \triangleq \int p(\underline{\eta})\ q(\underline{x}\ \underline{\eta})\ d\underline{\eta} \quad . \tag{31}$$

Using (15), this becomes

$$\pi(\underline{x}) = \int p(\underline{\eta})\ q\left(\underline{x} - \underline{f}_k(\underline{\eta})\right) d\underline{\eta} \quad . \tag{32}$$

We have renamed $\underline{x}_k$ as $\underline{x}$ and $\underline{x}_{k-1}$ as $\underline{\eta}$. The subscripts denoting the sampling time have been suppressed since they play no active role in the discussion. That is, the Bayesian recursion relations have the same form at every sampling time.

Next, the filtering density shall be rewritten in the following manner:

$$p(\underline{x}_k / \underline{z}^k) \triangleq p(\underline{x}) \triangleq c\pi(\underline{x})\ m\left(\underline{z} - \underline{h}_k(\underline{x})\right) \tag{33}$$

where c is the normalization constant, π is the prediction density as in (32), and m is the density of $\underline{z}$ given $\underline{x}$. The measurement $\underline{z}$ can be regarded as being known.

Consider the calculations required for one complete stage of the recursion. The filtering density p is computed as the product of π and m. Note that the calculations are started at t_1 with π equal to the a priori density $p(\underline{x}_1)$. Thus, $p(\underline{x}_k | \underline{z}^k) / c_k$ is readily formed for all k. The normalization constant is formed as

$$c^{-1} = \int \pi(\underline{x}) m\left(\underline{z} - \underline{h}(\underline{x})\right) dx \tag{34}$$

The integration generally must be accomplished numerically. It is immediately apparent that a considerable computational burden can be avoided if c is not determined. If one is interested in obtaining only the MAP estimator $\hat{\underline{x}}^{MAP}$, then c does not have to be found. However, if the mean-square estimator $\hat{\underline{x}}^{MS}$ is desired or if any moments of the distribution are to be computed, then an accurate value for c is required.

After determining p, the integrand in (32) can be formed. The prediction density π is obtained by carrying out the nonlinear convolution indicated in (32). Again, it is generally necessary to resort to numerical methods to determine π. Since π is a function of $\underline{x}$, the convolution implies that a large number of numerical integrals must be computed; essentially, an integration for each possible value of $\underline{x}$ is required.

After determining p and π, it is natural to compute moments of the a posteriori density. As noted above, the minimum mean-square estimate is provided by the conditional mean. The quality of the estimate is commonly gauged by forming the conditional covariance matrix. Conceivably, higher-order moments might also be determined as indicators of the effect of the nonlinearities and of the deviation of the a posteriori density from a gaussian. Of course, these are not ensemble statistics but are associated with a specific measurement realization.

A large number of methods for the evaluation of the Bayesian recursion relations have been proposed and studied. These methods have the common characteristic that the calculations are performed after defining a "grid." The grid points provide a finite collection on which approximations can be based. Obviously, these points are contained in a finite region of state space even though the integrations generally are carried out over infinite intervals. Thus, the functions must be such that there is negligible probability mass outside of the region containing the grid points. The manner in which the grid is defined is an important consideration in the development of an algorithm.

Let us consider an approach to the evaluation of the nonlinear convolution (32). Suppose that a specific value $\underline{x}_i$ is prescribed for $\underline{x}$ so the integration will yield a well-defined number. The numerical integration of (32), essentially requires that the integral be replaced by a summation involving a discretization of the integration-variable $\underline{\eta}$. The manner in which the grid points are defined may be accomplished arbitrarily or as in integral part of the quadrature method. For example, in an nth-order Gauss-Hermite quadrature, the grid points are chosen as the zeros of the nth Hermite polynomial. Let $\underline{\eta}_j$, $\{j = 1, 2, \ldots, N_{k-1}\}$ denote the N_{k-1} grid points for the variable $\underline{\eta}$. Furthermore, suppose that $\underline{x}_i$ is regarded as the ith grid point for the discretization of the

variable $\underline{x}$ into N_k points. Then, the convolution (32) is replaced by

$$p(\underline{x}_i) = \sum_{j=1}^{N_{k-1}} \alpha_j p(\underline{\eta}_j)\, p\left(\underline{x}_i - f(\underline{\eta}_j)\right) \; ; \quad i = 1,2,\ldots,N_k \; . \tag{35}$$

The coefficients α_i represent the weighting coefficients of the numerical integration scheme. Clearly, if there are a large number of grid points the storage and computational burden can be enormous, even for present-day digital computers.

Because of the storage and computational burden implied by solving the Bayesian recursion relations, it is natural to seek ways in which these requirements can be reduced. Effectively, the nonlinear filtering problem can be regarded in this completely computational context. In the subsequent discussion, we shall review some the approaches that have been proposed, summarize the types of results that have been obtained, and make suggestions for areas requiring additional investigation.

The earliest and by far the most extensively applied approach was motivated by the existence of the general solution for linear, gaussian systems (i.e., the Kalman filter). In this case, a single grid point is defined at each sampling time. Then, the system equations f and h are linearized relative to the grid point. This approximation of the system itself implies that the state and measurement perturbations are gaussian so the Kalman filter can be applied directly.

A number of generalizations to include higher order perturbations have been proposed. We shall discuss this class of methods in Section VI. Since the approximations can be regarded as being most accurate in some neighborhood of the single grid (or reference) point, we shall refer to them as local methods. More recently, a second class of techniques has emerged which explicitly attempt to obtain solutions by defining a grid over the entire region containing significant probability mass. This class shall be referred to as global methods and is discussed in Section VII.

VI. LOCAL NONLINEAR FILTERING METHODS

Virtually the only recursive nonlinear filtering method that has seen application to practical problems is the so-called extended Kalman filter. In this approach, a single grid point is defined at each stage and the system is linearized relative to this point. If the grid point is chosen as the "best" estimate (i.e., the approximation of the conditional mean), the resulting estimator is called the extended Kalman filter [2,10]. This is apparently the simplest possible approach since it involves a single grid point and linear equations at each sampling time. In addition, the grid point at the kth time is obtained directly from the previous grid point and the appropriate system equation. It is also a most crude approximation and its validity depends heavily on the quality of the linear approximation.

Practical experience has demonstrated that the assumptions inherent in the extended Kalman filter are often valid and satisfactory results are often obtained. There are also well-known disadvantages and difficulties associated with the application of the extended Kalman filter. The manifestation of these difficulties is commonly referred to as the divergence [2,10,16] problem. Divergence is said to occur when the actual error in the estimate becomes inconsistent with the error covariance matrix approximation provided by the filter equations. This situation arises because of errors in the filter model, either as a result of errors in the basic model or as a result of the linearization errors.

Experience with the extended Kalman filter in a variety of applications has led to the definition of a number of subproblems that may have to be solved in order to develop a useful algorithm.

A. Filter Initialization

Before utilizing the Kalman filter, it is often necessary to process a small amount of data to obtain reference values to be used in the linearizations. Regardless of the manner in which it is accomplished, the filter must be initialized with suitable values for the estimate and error covariance matrix in order to obtain reasonable estimates at subsequent times.

B. Form of the Filter Model

The linearization errors can be reduced in many cases by the form used for the system model. The choice of coordinate system can be important [17]. Furthermore, the use of transformations [18] to obtain models which are more easily linearized are often possible.

C. Iterative Calculations

To improve the linearizations, one can iterate through a small amount of the data (e.g., one sample at a time) and use improved estimates in the linearizations before reprocessing the data [10].

D. Divergence Control

Divergence often occurs because the model does not adequately describe the system. To compensate for model errors, the plant and measurement noise covariance matrices or the Kalman gain matrix directly can be increased. This has the effect of causing the error covariance matrix to be increased and in a way to cause past data to be discounted relative to more recent samples. A large number of methods have been devised to compensate for model errors [e.g., 2,10,19].

E. Second Order Filter

In our Bayesian context, the use of the extended Kalman filter implies that the a posteriori density is gaussian. This can be an extremely poor approximation of the actual density function if all

possible values of the state $\underline{x}$ are considered. Other local (or perturbative) schemes have been devised in an effort to improve the quality of the approximation. The obvious extension [20,21] is to consider retaining the second-order terms in the expansion of the system functions f and h. Commonly, the assumption is made that the a posteriori density is still gaussian even with the presence of the second-order terms. This assumption is made in order to overcome the "moment closure problem" which is discussed briefly below.

For the purposes of discussion, suppose that we are considering a scalar, second-order system

$$x_k = f_k x_{k-1} + g_k x_{k-1}^2 + w_{k-1} \quad , \tag{36}$$

$$z_k = h_k x_k + e_k x_k^2 + v_k \quad . \tag{37}$$

Suppose at the (k-1)th sampling time that we know

$$E\left[x_{k-1} | \underline{z}^{k-1}\right] = \hat{x}_{k-1|k-1} \quad , \tag{38}$$

and

$$\text{var}\left(x_{k-1} | \underline{z}^{k-1}\right) = p_{k-1|k-1} \quad . \tag{39}$$

The mean value $E\left[x_k | \underline{z}^{k-1}\right]$ is seen to be

$$E\left[x_k | \underline{z}^{k-1}\right] = \hat{x}_{k|k-1} = f_k \hat{x}_{k-1|k-1} + g_k\left(p_{k-1|k-1} + \hat{x}_{k-1|k-1}^2\right) \quad . \tag{40}$$

To determine the variance, we note that

$$x_k - E\left[\hat{x}_{k|k-1}\right] \triangleq \tilde{x}_{k|k-1}$$

$$= g_k \tilde{x}^2_{k-1|k-1} + \left(f_k + 2g_k \hat{x}_{k-1|k-1}\right)\tilde{x}_{k-1|k-1} - g_k p_{k-1|k-1} + w_{k-1} \qquad (41)$$

so

$$\text{var}\left(x_k | \underline{z}^{k-1}\right) = q_{k-1} + \left(f_k + 2g_k \hat{x}_{k-1|k-1}\right)^2 p_{k-1|k-1} \qquad (42)$$

$$+ 2g_k \left(f_k + 2g_k \hat{x}_{k-1|k-1}\right) \mu_{k-1|k-1}$$

$$+ g_k^2 \left(\nu_{k-1|k-1} - p^2_{k-1|k-1}\right) \quad ,$$

where $\mu_{k-1|k-1}$ and $\nu_{k-1|k-1}$ represent the third and fourth central moments of x_{k-1} given $\underline{z}^{k-1}$. Thus, the calculation of $\text{var}(x_k | \underline{z}^{k-1})$ requires knowledge of the first four central moments of x_{k-1} given $\underline{z}^{k-1}$. For this example, the calculation of the ith moments always requires knowledge of the 2ith moments at the preceding time. This implies that one must know moments of every order and is referred to in general as the moment closure problem. To close the problem, it is common to assume that moments of order greater than some integer correspond with gaussian moments. For example, if the 3rd and higher order moments are assumed to be gaussian, then for all k

$$\nu_k = 0 \quad , \qquad (43)$$

$$\mu_k = 3p^2 \quad . \qquad (44)$$

F. Higher Order Polynomial Expansion

The first serious attempt to eliminate the gaussian assumption involved the use of Gram-Charlier or Edgeworth expansion [22]. The expansion is a series of polynomials which are orthogonal with respect to a gaussian distribution and can be used to represent a wide class of density functions. The initial use of this non-gaussian approximation was based on a perturbative approximation. As a consequence, it suffered from the disadvantage that a large number of terms were required to obtain a reasonable approximation of a distinctly nongaussian density. The behavior of the estimator obtained from this density approximation was found to be very sensitive to the quality of the approximation. When the infinite series is truncated, as it must for practical application, the resulting series can become negative over portions of the state space. Consequently, the density approximation is not itself a density. This can introduce unexpected influences into the behavior of the estimator, particularly if the integral over the region in which the function is nonpositive has a nontrivial value. Subsequently, other density approximations using the Edgeworth expansion have been proposed [e.g., 23,24,34]. This local method seems to be most useful when the a posteriori density is unimodal even though it is not Gaussian.

G. Parameter Identification

In certain cases the dynamic system, the measurement function, or even statistics of the noise or initial state can contain unknown and constant elements. Much of the work classified as system identification addresses this problem which is a very special case of the nonlinear estimation problem. Good descriptions of these techniques are contained in references [25,26].

VIII. GLOBAL NONLINEAR FILTERING METHODS

The obvious disadvantage of the local methods stems from the use of a single grid point on which to base the approximation. During the past few years, several methods have been proposed which attempt to improve the approximation by considering the density at many points selected through the region containing nonnegligible probability mass. These methods can be regarded as representing specific examples of ways in which the numerical integrations discussed in Section V can be accomplished. Some of these global approximations are reviewed in this section.

Quite possibly, the first step toward the development of a global method was taken by Magill [27] with a subsequent generalization by Hilborn and Lainiotis [28]. They considered linear systems with unknown parameters. To deal with this nonlinear problem,

a grid was established by discretizing the unknown parameters and by considering the resulting collection of linear filtering problems. A global method for the general nonlinear filtering problem was proposed by Bucy [29] when he introduced the point-mass method. This approach was elaborated upon by Bucy and Senne [30] at the First Symposium on Nonlinear Estimation in 1970. At this same meeting Alspach and Sorenson [31] proposed the gaussian sum approximation as an alternative approach. Subsequent Symposia on Nonlinear Estimation included many extensions and saw the introduction of other techniques. Center [32] provided a unifying theoretical framework by considering the problem in the context of generalized least-squares. His approach permits, conceptually at least, the development of a countless number of approximations. In the Second Symposium on Nonlinear Estimation, Center discussed as specific examples the point-mass, gaussian sum, and Edgeworth expansion approximations. Later [33], he also discussed the spline approximation method proposed by Jan and de Figueiredo [12].

All specific global methods must provide solutions of the following general problems.

(a) An initial grid must be defined. It is important that the region encompassed by the grid includes the true value of the state. In addition, the number and manner in which the grid points are distributed within the approximation region must be defined.

(b) A procedure must be defined for defining the grid at each subsequent sampling time. While the grid could be the same throughout, the dynamic nature of the problem and the desire for computational efficiency indicate the advisability of redefining the grid at each sampling time.

(c) Given the grid, a method must be selected for approximating the functions and/or for carrying out the Bayes' rule calculations. The approximation method and the grid selection method are not unrelated and the implementation of a particular method may require interaction between the two considerations.

Below we will show in some detail how much interaction occurs for one of the methods in order to give the reader a feel for such interaction in one particular case. The other methods have been applied in similar problems and are briefly described below and described in detail in the references.

Alspach and Sorenson [11,31,34,35] proposed approximating the a posteriori density function by a weighted sum of Gaussian densities. For example, a density p is approximated by the density* p_a

$$p_a(\underline{x}) = \sum_{i=1}^{q} \alpha_i N_x(\hat{\underline{x}}_i, P_i) , \tag{45}$$

where the weighting coefficients α_i are nonnegative and $\sum_{i=1}^{q} \alpha_i = 1$.

*$N_x(\underline{a},B) \triangleq (2\pi)^{-n/2}(\det B)^{-1/2} \exp\{-\frac{1}{2}(\underline{x} - \underline{a})^T B^{-1} (\underline{x} - \underline{a})\}$.

This approximation is motivated by the realization that p_A converges uniformly to p for a large class of densities. Thus, the approximation p_A can be made as accurate as one wants through the choice of q. The idea of using this type of approximation has been suggested by several others [e.g., 13,14,15,36].

After q has been defined, it is necessary to assign values to the parameters α_i, $\hat{\underline{x}}_i$, P_i, {i = 1, 2, ..., q}. The mean values $\hat{\underline{x}}_i$ represent grid points for the approximation. The selection of all of these parameters must yield a satisfactory representation of the a posteriori density. It is natural to formulate their determination as an optimization problem. Let us choose α_i, $\hat{\underline{x}}_i$, P_i, {i = 1, 2, ..., q} so that the generalized least-squares performance index,

$$\epsilon_{LS} = \int [p(\underline{x}) - p_a(\underline{x})]^2 \, d\underline{x}, \tag{46}$$

is minimized subject to the constraints that for all i, $\alpha_i \geq 0$, $\Sigma \alpha_i = 1$ and P_i is a positive semidefinite matrix. Figures 2 to 4 show the approximations resulting in fitting three different scalar densities by such a direct optimization method. These densities contain most of the features that can give difficulty in the various density functions encountered in practice. These difficulties include discontinuities, skewness, unboundedness, and the problem of converging to zero both faster and slower than the gaussian.

The first example that is discussed here is the gamma density function. It is defined as:

$$p(x) = \begin{cases} 0 & \text{for } x < 0 \\ \dfrac{x^3 e^{-x}}{6} & \text{for } x \geq 0 \end{cases} \tag{47}$$

This distribution has a mean value of 4 and second, third, and fourth central moments of 4, 8, and 72, respectively. Figure 2 shows the result of fitting one to four gaussians to this density. Note that in two cases several moments of the approximate gaussian sum density were constrained to match the moments of the true density. The bad effect on the L_2 fit indicates difficulties with this and other moment matching techniques.

The second density approximated is a uniform density function:

$$p(x) = \begin{cases} \frac{1}{4} & -2<x<2 \\ 0 & \text{elsewhere} \end{cases} \tag{48}$$

The obvious symmetry was imposed on all approximate densities in order to exactly match all odd moments. The results of fitting 2 to 5 gaussians to this density are shown in Figure 3. Note the appearance of a "Gibbs phenomenon" on the last approximate density.

The last density approximated and reported here is a product of two independent zero mean gaussian random variables.

$$z = xy \tag{49}$$

$$p(x) = N(x, 4) \tag{50}$$

$$p(y) = N(x, 4) \tag{51}$$

$$p(z) = K_0 (z/4) / 4\pi \tag{52}$$

where K_0 is the modified bessel function of the second kind of order zero. Because of the symmetry of this density, all odd moments are zero. The second and fourth central moments are:

$$\sigma^2 = E(z^2) = 16 \quad \text{and} \quad E(z^4) = 2304 \tag{53}$$

This density and one, three, and five gaussian sum approximations to it are shown in Figure 4.

The development of such approximations requires the utilization of numerical search techniques. The approximations can be determined off line but may require extensive calculation. This approach may not be acceptable in many circumstances. Thus, in reference [34], an alternate sampler method was developed and entitled the "theorem fit" approach. This is done as follows:

1. The number of Gaussian terms in the sum, n, is chosen based on previous experience.
2. The region (a,b) over which the density function is to be approximated is chosen.
3. The mean values a_i for each Gaussian are placed uniformly inside (a,b).

4. The weighting functions α_i are selected to be proportional to the value of the density function to be approximated at a_i, and are normalized to one.

5. The value of the variance is taken to be independent of i and is found either by:

   5.1 By a one-dimensional search to minimize the L_2 error;

   or

   5.2 Is chosen such that, $\sigma = z \frac{(b-a)}{n}$, for a prespecified value of z.

The first approach is called the "best theorem fit" and the second the "smoothed theorem fit"; no search at all is required in this second method of obtaining an approximate density.

If J involves more than one region, a modification of this technique has been used which treats each region separately and takes the number of terms in each region to be proportional to its measure. Such approximate densities for the uniform and gamma densities are shown in Figures 5 and 6. In these figures the parameter z has been chosen to be .6.

Gaussian sum densities can also be utilized to approximate densities of greater than one dimension. In doing this it is possible to take into account natural symmetries of the density to be approximated. For example, suppose the measurement function is given by:

$$z = h(\underline{x}) = .1\, x_1 + (x_1 - x_2)^2 + v_k \tag{54}$$

$$E\,(v_k^2) = .1 \tag{55}$$

The measurement function is then given by

$$P(z|\underline{x}_1) = N(z - .1x_1 + (x_1 - x_2)^2,\ .1) \tag{56}$$

or for the particular case of z equal to 1 this function is shown in Figure 7(a). If the a priori mean $\hat{\underline{x}}$ is zero

$$\hat{\underline{x}} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{57}$$

and this is taken as the linearization point, the approximation utilized in an extended Kalman filter for this function is shown in Figure 7(b). A simple two-term gaussian sum shown in Figure 7(c) gives a far better approximation to the true density and a 30-term smoothed theorem fit density shown in Figure 7(d) captures even more of the fine details of the original function.

Another example of using a gaussian sum density to approximate a two-dimensional function arises in the passive bearing's only tracking problem (reference [34]). A target is located at $\underline{x}_k$ or $(x_k, y_k)^*$ in Figure 8(a) and is observed by a ship at location S which measures the angle α. This geometry is shown in Figure 8(a). The measurement function is

$$z = h(\underline{x}_k) = \tan^{-1}\left[(y_k - \sin\beta_k)/(x_k - \cos\beta_k)\right] + v_k \qquad (58)$$

$$E(v_k^2) = \sigma_v^2 = .01 \text{ radian}^2 \qquad (59)$$

and the function $P(z_k/\underline{x}_k)$ is shown in Figure 8(b). If an extended Kalman filter linearizes this function around the most recent line of bearing the approximate function shown in Figure 8(c) results. This is too wide if the target is closer than the a priori estimate and too narrow if the target is farther from the origin than the original estimate. This feature of the approximation can lead to "range collapse" or divergence where the estimate steps to the origin. However, the general shape of this extended Kalman filter approximation is correct. If one linearizes about the last estimated position which, however, happens not to lie in the measured bearing, very bad approximations, away from the linearization point, can result. Then one gets functions which bear little resemblance to the true measurement function just as in the last example. A ten-term gaussian sum "smoothed theorem fit" is shown in Figure 8(d). Note in Figures 8(b) and 8(d) it has been assumed that the true state cannot lie greater than six orbital radii away from the observer. This accounts for the cutoffs on both figures.

There are a variety of ways in which these general approximations can be utilized. For example, if the a priori density is approximated by a Gaussian sum density with N terms, then this defines a generalized grid in the initial state space. If the plant

and measurement functions are linear with gaussian perturbing noise, then the evolution of each term in the a priori gaussian sum density is described by a linear Kalman filter. For example, in the simple example described earlier with a scalar state and uniform initial state density, the evolution of the true density, the gaussian sum approximate density, and a single linear Kalman filter density with time are shown in Figure 9.

In this case only the a priori density $P(\underline{x}_1) = \pi(\underline{x})$ needs to be approximated by a gaussian sum, and the a posteriori density $P(x_k/\underline{z}^k)$ can easily be shown to be a gaussian sum with N terms.

The more general case of non-gaussian plant and measurement noise each approximated by gaussian sum has also been considered. In this case the a posteriori density is also a gaussian sum but the number of terms in the density grows with the number of stages. This is shown in detail in reference [11].

The more general case of a nonlinear measurement equation ($z_k = h_k(\underline{x}_k)$) is considered in reference [35]. Here, in order to enfold the measurement z, it is necessary to multiply $\pi(x)$ by $m(z-h(\underline{x}))$. If h is nonlinear, the product is no longer a Gaussian sum. One approximation at this point is to linearize h(x) about each of these grid points. Because the variance of each term of the sum is small, the linearization must be valid only in a small region surrounding the grid point. The extended Kalman filter can be

applied at each grid point to obtain the means $\hat{\underline{x}}_i$ and covariances P_i needed in the gaussian sum approximation of the filtering density from that of the prediction density. The extended Kalman filter is used to obtain a grid for the next sampling time and to obtain the gaussian sum approximation of the prediction density π. Next, using Eq. 32, the new predicted density $\pi(x)$ is calculated. If the system dynamics are linear, then this can be solved exactly and again the predicted density is a Gaussian sum. If not, $f_k(\underline{x})$ must be linearized about each grid point as in an extended Kalman filter and then the next stage density is again in a Gaussian sum form. A simple quadratic scalar example of this is taken from reference [35].

Consider the scalar system with the plant described by

$$x_{k+1} = x_k + \eta x_k^2 + w_k. \tag{60}$$

The state x_k is to be estimated from the measurement data $\underline{z}^k$ where

$$z_k = x_k^2 + v_k, \qquad k = 0,1,\cdots . \tag{61}$$

The initial state and the plant and measurement noise sequences are independent, white, and gaussian with

$$E(\hat{x}_0') = 1; \quad E[(x_0 - \hat{x}_0')^2] = 1 ; \tag{62}$$

$$E(w_k) = E(v_k) = 0 ; \tag{63}$$

$$E(w_k^2) = \sigma_w^2 ; \qquad E(v_k^2) = \sigma_v^2 . \tag{64}$$

The a priori mean and variance of the initial state are held at these values for all examples presented here, although others have been investigated. The basic parameters of the system in the present study are the variances of the plant and measurement noise and the relative effect of the plant nonlinearity η. These variances have been chosen to be independent of k for clarity of presentation only. The value of each of these parameters will be specified for each case presented.

Results for four different filters are presented and discussed, although not all results are included in Figures 10 and 11. When a filter performs very badly, it may fall off the scale of the charts and thus not be shown. The first three are filters that have been considered previously in the literature and in which the a posteriori and predicted density functions are assumed to be gaussian at each stage. The first of these is the extended Kalman filter. This is the filter most often used in practice. The second filter uses one iteration to improve the reference values used in the linearization. The third filter is the gaussian filter of [20], where second-order terms are used to modify the mean and variance of the next stage predicted and a posteriori density functions. The fourth is the gaussian sum filter.

The characteristics of the filtering problem depend heavily on the position of the state variable x_k with respect to the point of symmetry of the measurement nonlinearity. When x_k is near that point (zero in this case), the ratio x_k^2/σ_v is small and the gaussian filters tend to diverge. As the state moves away from this point, the measurement nonlinearity becomes increasingly more negligible and the gaussian filters tend to perform well. This is particularly clear when there is no plant nonlinearity $\eta = 0$ and no plant noise $\sigma_w = 0$. In this case the relative performance of the different filters depends most strongly on the value of the state variable and less on the particular measurement realization under consideration. For this reason it was found best with a limited number of realizations to choose the true initial value of state as a parameter and only select the measurement and plant noise from a random number generator. This was particularly useful in the Monte-Carlo averages, but was done in all the cases presented below.

When there is no plant nonlinearity [$\eta = 0$ in (60)], it is impossible from the available measurement data to discriminate between the true value of the state and the negative of that value. Thus $p(x_k|z^k)$ should become bimodal if the value of the state is nonzero. This is, of course, not possible for any of the gaussian filters. When there is no plant noise or nonlinearity, the a posteriori density can be computed exactly. Under these conditions it is (except for a normalization constant) simply given by

$$P(x_k|z^k) = cP_{x_0}(x_0)\, P_{v_0}(z_0 - h(x_0)) \cdots P_{v_k}(z_k - h(x_k)) \quad . \quad (65)$$

The density function of a specific realization is depicted in Figure 10. The values of the system parameters are stated in the figure. The gaussian sum filter provided an approximation that is indistinguishable from the true a posteriori density for the example. In this case the a priori density $p(x_0)$ was approximated by a sum of 40 gaussians. Observe that the second-order filter provides an extremely conservative result and estimates the state to be zero instead of ±0.2. The extended Kalman filter tends to diverge. Only the iterated filter performs at all satisfactorily and provides an estimate of approximately 0.2.

It is interesting that the minimum variance estimate that one would obtain from $p(x_k|z^k)$ provides an estimate that is between the two peaks (i.e., since the conditional mean is the minimum variance estimate). Clearly, this estimate is very conservative and, consequently, may be unsatisfactory. A maximum likelihood estimate would yield a value close to the true value or its negative.

When a plant nonlinearity from (61) is included, it is possible to distinguish between the two values and the gaussian sum filter quickly selects the proper peak. This is shown in Figure 11 where the value of η is -0.2. Since the state has a negative value, the gaussian filters all perform unsatisfactorily, so only the results

of the gaussian sum filter are shown. This example demonstrates the difficulty that a maximum likelihood estimator might encounter. It is observed that the maximum value of $p(x_k|z^k)$ switches back and forth from positive to negative. Without complete knowledge of the density function, it is unlikely that a procedure could be devised that would reflect this behavior.

In the previously described use of gaussian sums, the gaussian sum approximation takes the form of a number of extended Kalman filters operating in parallel. It is easy to obtain an indication of the computational burden that is associated with this nonlinear filter. If q extended Kalman filters are required at each stage of the sequence, then the gaussian sum requires approximately q times as much effort as a single filter. The burden of a single filter is well known [e.g., 37]. The general use of parallel processors in this problem has been considered by several authors [38 - 41].

In some problems it makes more sense to approximate $P(\underline{x}_k/\underline{x}_{k+1})$ or $P(\underline{z}_k/\underline{x}_k)$ directly as a gaussian sum at each stage. In this case, the number of terms in the gaussian sum approximation to the a posteriori density grows with the number of stages. It is, however, possible to drop any term whose weighting coefficient, α_i, is negligible. It is also often possible to combine two or more terms whose grid points from prediction fall sufficiently close together. In this way the total number of terms in the gaussian sum can be controlled.

An example of this type is the vector tracking example whose measurement density function $P(z_k/x_k)$ was described earlier in Figure 8 and Equation (58).

The state vector propagates according to the linear plant

$$\underline{x}_{k+1} = \underline{x}_k + \underline{w}_k \qquad \underline{x}_k = \begin{pmatrix} x_k \\ y_k \end{pmatrix}$$

and the state is observed by the scalar nonlinear measurement function of Equation (58) where

$$\beta_k = \beta_0 + \dot{\beta}\,(k - 1)$$

where β_0 and $\dot{\beta}$ are given constants. The a priori random variables $\underline{x}_0$, v_k, and $\underline{w}_k$ are white, independent, gaussian random variables and sequences.

The preceding model arises in connection with the tracking geometry of Figure 8 where target T at the position defined $\underline{x}_k^T = (x_k\ y_k)$ is undergoing a random walk in the two-dimensional state space. The observer S is passively measuring the line-of-sight α as it travels in a deterministic orbit around the unit circle.

Results obtained from the application of the gaussian sum filter to a specific example are shown in Figure 12. The position of the observer is shown by the cross on the unit orbit and the

cross on the density function shows the true position of the target. The *a priori* estimate for the initial state was taken to be

$$\hat{x}_0' = \begin{pmatrix} 2 \\ -0.2 \end{pmatrix} \qquad P_0' = \begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix},$$

while the true value of the initial state (and all subsequent values since there is no plant noise) was taken to be

$$x_k = \begin{pmatrix} 0 \\ 0.5 \end{pmatrix}.$$

The measurement noise has a one sigma value of 0.01 rad or about one-half degree. The non-gaussian *a posteriori* filtering-density function is seen to propagate from stage 1 to stage 9 in this figure where a measurement is taken every 10°.

In Figure 13 results obtained using the extended Kalman filter and the gaussian sum filter are compared. The parameters ϵ_{xk}, ϵ_{yk}, and A_k for a single realization are presented. The improvement provided by the gaussian sum filter is striking.

A number of other interesting nonlinear non-gaussian stochastic dynamic systems have been investigated utilizing these techniques in the literature, and the reader is directed there for more details on specific problems.

If the covariance term in the "theorem fit" method of approximating densities is made very small, one ends up with gaussian shaped delta functions with a weighting equal to the function value at that point. Such an approximation is a very bad fit to the density in an L_n sense but distributions and moments for such an approximation can be arbitrarily close to those from the exact density as the density of the grid increases. In this way one can move from the gaussian sum density approximation to the "point mass" approximation.

The point mass approximation was introduced by Bucy and Senne in 1970. Bucy [29] and Bucy and Senne [30] have suggested that the error covariance matrix be used to establish the region and the grid. Essentially, the eigenvectors are used to define the principal axes. The grid is centered at the mean value. The grid along each axis was chosen to extend over a distance sufficient (e.g., 16 times the magnitude of the corresponding eigenvalue) to insure that the true state is contained within the grid region. The number of grid points is prescribed to provide an adequate approximation. The basic method of defining the grid is modified to suit the requirements imposed by a particular problem. For example, when the a posteriori density is multimodal, it is reasonable to define a grid for each mode rather than for the entire density.

The manner in which the grid is updated at the next sampling time is straightforward since the system dynamics provide the mean values and the covariance matrix for the predictor density.

Once the grid points have been established, the density functions can be evaluated at each of them. These values, after being suitably normalized, can be regarded as point masses for a discrete approximation of the distributions. Using the point-mass approximation, the Bayesian recursion relations are readily evaluated. This approximation is essentially equivalent to using a rectangular integration rule to accomplish the numerical quadratures.

IX. NONLINEAR FILTERING--A CRITICAL LOOK

Global nonlinear filtering is growing beyond its infancy. As must be true for any infant, the first steps, as exemplified by the work mentioned above, are exhilarating for those involved and can easily lead to overly ambitious claims and unwarranted optimism. Viewed with even a modicum of perspective, however, it becomes obvious that much work remains before the infant will grow to maturity. It is fun and, hopefully, worthwhile to attempt to predict the character of the mature development and to suggest some activities that are required to shape the development.

The basic objective of global nonlinear filtering might be regarded as the development of a practical computational algorithm which will permit the determination of the a posteriori density to any prescribed accuracy for any system. This is the achievement of the Kalman filter for linear, gaussian systems; if it can be accomplished for nonlinear non-gaussian systems, the achievement would be worthy of any of the scientific titans of history. The developments described above do provide procedures for computing the a posteriori density for any system. But they have the practical limitation that the computational requirements associated with their implementation are enormous. Thus, the development of an algorithm must be guided by the requirement of achieving computational efficiency. With the rapid development of mini-computers, it appears that practical nonlinear filtering may be possible using special-purpose rather than general-purpose digital computers. It appears reasonable to consider, for example, the use of mini-computers for parallel processing. Possibly, some of the general ideas discussed by Korn [42] will prove useful.

Assuming that global nonlinear filtering methods will continue to require substantially more computation than local filtering techniques, it is natural to ask and attempt to answer the following questions. Under what conditions is it desirable and necessary to assume the additional computational burden and utilize global nonlinear filtering techniques? Certainly, no answer to this

question that could be universally accepted exists at this time. However, some related considerations can be discussed.

Local filtering techniques in general and the extended Kalman filter (EKF) in particular are looked upon with scorn in some quarters because these approaches are "suboptimal." In addition, the degree of suboptimality is not readily determined. As a consequence, it is reasonable to solve the global filtering problem if only to provide a reference against which local methods can be compared. However, the continued use of the EKF must be tolerated because it has proven to provide satisfactory results for many nonlinear systems. This is especially true when the filter is designed to monitor the residuals and to initiate corrective action whenever a low frequency component is observed that implies the onset of divergence.

The success of the EKF forces a search for general circumstances in which this local filtering method cannot be expected to perform satisfactorily. Certainly, one of the most important requirements is that an a priori estimate be available which permits the local approximation to be valid initially. If it is impossible to define an appropriate a priori estimate, then the EKF is doomed to failure and a global filter is required. For many systems of interest, this would appear to be an unlikely situation. Frequently, the signal-to-noise ratio is sufficiently large that a reasonable estimate can be obtained using only deterministic models. When

more than one solution is possible, physical consideration may permit the determination of the only reasonable solution which can then be used to initiate the EKF. If more than one a priori estimate must be considered, the a posteriori density will be multimodal so the EKF cannot be used.

If the a posteriori density can be regarded as unimodal but non-gaussian, the EKF must produce suboptimal results. Thus, it may be desirable to utilize local or global procedures which eliminate the gaussian assumption. In many cases, the EKF can be expected to provide pessimistic results since the gaussian density maximizes entropy. As long as the residual is forced to be white, the EKF should produce results that are satisfactory in some ways. More complicated procedures may provide improvements but this would seem to be very problem-dependent.

Finally, the signal-to-noise ratio may be so small that linearizations provide inadequate approximations with the result that the EKF produces little data filtering. That is, the divergence control logic may require past data to be discounted so strongly that only current data is used in determining the estimate. Then, the estimation error will be comparable or greater than the measurement noise indicating the lack of any filtering (noise removal) activity. In this case, global nonlinear filters may be required in order to extract the maximum information from the data.

Among the advantages that can result from the use of global nonlinear filtering are the following:

(a) It is not necessary to have *a priori* estimates of the state that are sufficiently accurate to validate the linearization. Thus, the problem of initializing the filter is eliminated.

(b) Situations in which the *a posteriori* density is multimodal are handled in a straightforward manner. The consideration of multimodality enters primarily through the definition of the grid and the choice of estimator criterion.

(c) The elimination of the assumption that the *a posteriori* density is gaussian can permit more accurate statistical statements to be made. A simple example is given in Ref. [19] which demonstrates the insights possible from knowledge of the *a posteriori* density.

(d) Calculation of the *a posteriori* density provides a meaningful reference which can be used to measure the performance of all suboptimal procedures. The accurate calculation of $p(x_k/z^k)$ permits one to more rationally evaluate the effects of the approximations used in suboptimal estimators. Generally, even suboptimal estimators approach the optimal response of the global filter after a large quantity of data has been processed. The difference in transient response can be determined and can provide a measure of the adequacy of a particular suboptimal algorithm.

In the study of nonlinear filtering, it is not surprising to find that there are few analytical results and closed-form solutions. Thus, to deal with these problems, it is natural to see a concentration of effort on the development of computational procedures. In this sense, the field is similar to the study of nonlinear programming. Unlike the latter, we do not have standard test problems nor extensive numerical studies of different algorithms which have been developed for the same general problem. It seems that this is a gap that must be filled. Several problems that have appeared in the literature and have been described above can serve as candidates for standard test problems. Rational criteria for comparing algorithms need to be established. It should be incumbent upon the proposer of a new algorithm to provide meaningful comparisons of his procedure with existing algorithms. By this means one can hope to establish situations in which specific algorithms will have demonstrable advantages.

As nonlinear filtering begins to see practical application, a wealth of new problems will be uncovered and the research will progress into new areas. A question which requires immediate consideration arises when we contemplate the basic assumptions implicit in the Bayesian recursion relations. This solution of the nonlinear filtering problem supposes that we have a complete probabilistic description of the system. In practice, one often considers himself lucky to have information about the second moments. Thus, it is naive to believe that the probabilistic model is justified.

Consequently, it is imperative that the sensitivity to model errors be examined in considerable detail. On one hand, it might be possible to reduce the computational burden associated with the current global filters by exploiting the knowledge that model errors exist. On the other hand, sensitivity to model errors might indicate the folly of the Bayesian approach entirely and cause the redirection of research activities into less model-dependent formulations.

REFERENCES

1. R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," Trans. ASME:J. Basic Engr., 82D, 1960, pp. 34-45.
2. H. W. Sorenson, "Kalman Filtering Techniques" in Advances in Control Systems, 3 (edited by C. T. Leondes), Academic Press, New York, 1966.
3. R. Bellman, R. Kalaba, "A Mathematical Theory of Control Adaptive Processes," Proc. Nat'l Acad. Sci., 45, 1959, pp. 1288-1290.
4. Y. C. Ho, R. C. K. Lee, "A Bayesian Approach to Problems in Stochastic Estimation and Control," Proceedings of the JACC, 1964, pp. 382-387.
5. R. C. K. Lee, Optimal Estimation, Identification, and Control, Research Monograph No. 28, MIT Press, Cambridge, Massachusetts, 1964.
6. R. Deutsch, Estimation Theory, Prentice-Hall, Englewood Cliffs, N.J., 1965.
7. M. Aoki, Optimization of Stochastic Systems, Academic Press, New York, 1967.
8. A. A. Feldbaum, Optimal Control Systems, Academic Press, New York, 1965.
9. A. E. Bryson, D. E. Johansen, "Linear Filtering for Time-varying Systems Using Measurements Containing Colored Noise," IEEE Trans. Auto. Cont., AC-10, 1965, pp. 4-10.
10. A. H. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, New York, 1970.

11. H. W. Sorenson, D. L. Alspach, "Recursive Bayesian Estimation Using Gaussian Sums," Automatica, 7, 1971, pp. 465-479.

12. R. J. P. de Figueiredo, Y. Jan, "Spline Filters," Proc. 2nd Symp. Nonlinear Estimation, San Diego, Western Periodicals, North Hollywood, CA, 1971.

13. A. V. Cameron, "Control and Estimation of Linear Systems with Non-Gaussian a priori Distributions," Proc. Sixth Ann. Allerton Conf. on Circuits and Systems, Urbana, Ill., 1968.

14. P. J. Buxbaum, R. A. Haddad, "Recursive Optimal Estimation for a Class of Non-Gaussian Processes," Proc. Symp. Comp. Comm., Poly. Inst. of Brooklyn, 1969, pp. 375-399.

15. J. T. Lo, "Finite-dimensional Sensor Orbits and Optimal Nonlinear Filtering," IEEE Trans. Infor. Theory, IT-18, 1972.

16. R. J. Fitzgerald, "Divergence of the Kalman Filter," IEEE Trans. Auto. Cont., AC-16, 1971, pp. 736-747.

17. R. K. Mehra, "A Comparison of Two Nonlinear Filters for Ballistic Trajectory Estimation," Proc. First Symp. Nonlinear Estimation, San Diego, California, Western Periodicals, North Hollywood, CA, 1970, pp. 278-281.

18. D. W. Whitcombe, "Pseudo State Measurements Applied to Recursive Nonlinear Filtering," Proc. Third Symp. Nonlinear Estimation, San Diego, California, Western Periodicals, North Hollywood, CA, 1972, pp. 278-281.

19. H. W. Sorenson, J. E. Sacks, "Recursive Fading Memory Filtering," Information Sciences, 3, 1971, pp. 101-119.

20. H. W. Sorenson, A. R. Stubberud, "Recursive Filtering for Systems with Small but Non-Negligible Nonlinearities," Int. J. Cont., 7, 1968, pp. 271-280.

21. M. Athaus, R. P. Wishner, A. Bertolini, "Suboptimal State Estimators for Continuous Time Nonlinear Systems from Discrete Noisy Measurements," IEEE Trans. Auto. Cont., AC-13, 1968, pp. 504-514.

22. H. W. Sorenson, A. R. Stubberud, "Nonlinear Filtering by Approximation of the A Posteriori Density," Int. J. Cont., 8, 1968, pp. 33-51.

23. W. Kizner, "Optimal Nonlinear Estimation Based on Orthogonal Expansion," JPL Tech. Report, 32-1366, 1969.

24. K. Srinivasan, "State Estimation by Orthogonal Expansion of Probability Distributions," IEEE Trans. Auto. Cont., AC-15, 1970, pp. 3-10.

25. D. G. Graupe, Identification of Systems. Van Nostrand-Rheinhold, New York, 1972.

26. P. Eykhoff, System Identification: Parameter and State Estimation, John Wiley, New York, 1974.

27. D. T. Magill, "Optimal Adaptive Estimation of Sampled Stochastic Processes," IEEE Trans. Auto. Cont., AC-10, 1965, pp. 434-439.

28. C. G. Hilborn, D. G. Lainiotis, "Optimal Estimation in the Presence of Unknown Parameters," IEEE Trans. Sys. Sci. Cyb., SSC-5, 1969, pp. 38-43.

29. R. S. Bucy, "Bayes' Theorem and Digital Realization for Nonlinear Filters," J. Astronaut, Sci. 17, 1969, pp. 80-94.

30. R. S. Bucy, and K. D. Senne, "Realization of Optimum Discrete-Time Nonlinear Estimators," First Symp. on Nonlinear Est., San Diego, California, Western Periodicals, North Hollywood, CA, 1970, pp. 6-17.

31. D. L. Alspach, and H. W. Sorenson, "Approximation of Density Function by a Sum of Gaussians for Nonlinear Bayesian Estimation," First Symp. on Nonlinear Est., San Diego, California, Western Periodicals, North Hollywood, CA, 1970, pp. 19-31.

32. J. L. Center, "Practical Nonlinear Filtering of Discrete Observations by Generalized Least-Squares Approximation of the Conditional Probability Distribution," Second Symp. on Nonlinear Est., San Diego, California, Western Periodicals, North Hollywood, CA, 1971, pp. 88-99.

33. J. L. Center, "Practical Nonlinear Filtering Based on Generalized Least-Squares Approximation of the Conditional Probability Distribution," Ph.D. dissertation, Washington University, St. Louis, 1972.

34. D. L. Alspach, "A Bayesian Approximation Technique for Estimation and Control of Time Discrete Stochastic Systems," Ph.D. dissertation, University of California, San Diego, California, 1969.

35. D. L. Alspach, H. W. Sorenson, "Nonlinear Bayesian Estimation Using Gaussian Sum Approximations," IEEE Trans. Auto. Cont., AC-17, 1972, pp. 439-448.

36. D. G. Lainiotis, "Optimal Nonlinear Estimation," Int. J. Cont., 14, 1971.

37. J. M. Mendel, Computational Requirements for a Discrete Kalman Filter, IEEE Trans. Auto. Cont., AC-16, 1971, pp. 748-758.

38. E. Tse, Parallel Computation of the Conditional Mean State Estimate for Nonlinear Systems, Second Symp. Nonlinear Est., San Diego, California, Western Periodicals, North Hollywood, CA, 1971, pp. 385-394.

39. J. E. Shore, Second Thoughts on Parallel Processing, Computers Elect. Engr. 1, 1973, pp. 95-110.

40. C. Hecht, Digital Realization of Nonlinear Filters, Second Symp. on Nonlinear Est., San Diego, California, Western Periodicals, North Hollywood, CA, 1971, pp. 152-158.

41. R. S. Bucy, M. J. Merritt, and D. S. Miller, Hybrid Computer Synthesis of Optimal Discrete Nonlinear Filters, Second Symp. on Nonlinear Est., San Diego, California, Western Periodicals, North Hollywood, CA, 1971, pp. 59 87.

42. G. A. Korn, "Back to Parallel Computation: Proposal for a Completely New On-Line Simulation System Using Standard Minicomputers for Low-Cost Multiprocessing," Simulation, 19, 1972, pp. 37-46.

FIGURE LEGENDS

Figure 1. The true value of one sigma and the Kalman approximation to it versus z_0.

Figure 2. One to four Gaussians in L^2 search fit to a gamma density.

Figure 3. Gaussian sum approximation to uniform density L^2 search fit 2 to 5 terms.

Figure 4. 1, 3, and 5 L^2 search fit approximations to product density.

Figure 5. Gaussian sum approximations of uniform density functions.

Figure 6. 6 and 10 term Gaussian sum theorem fit approximation to a gamma density.

Figure 7. Measurement density function and approximation.

Figure 8. The passive, bearings-only tracking problem.

Figure 9. Behavior of the a posteriori density--true and approximate.

Figure 10. Filtering density and approximations. Solid line is true PDF. Broken line is Gaussian sum, $\underline{x} \cdots \underline{x}$ is second order. $+ \cdots +$ is iterated.

Figure 11. Gaussian sum approximation to filtering density for nonlinear plant and measurement. Solid line is Gaussian sum PDF. $o \cdots \bullet$ is true value of state. $\underline{x}_0 = -0.2$, $\eta = -0.2$, $\sigma_{\underline{w}} = 0$, and $\sigma_{\underline{v}} = 0.05$.

Figure 12. Filtering density for vector tracking example.

Figure 13. Relative performance of extended Kalman and Gaussian sum filters for tracking problem. Broken line denotes Kalman. Solid line denotes Gaussian sum.

FIGURE 1

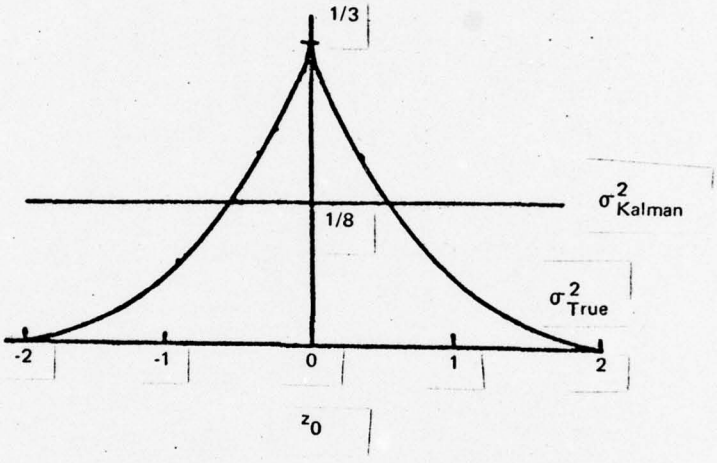
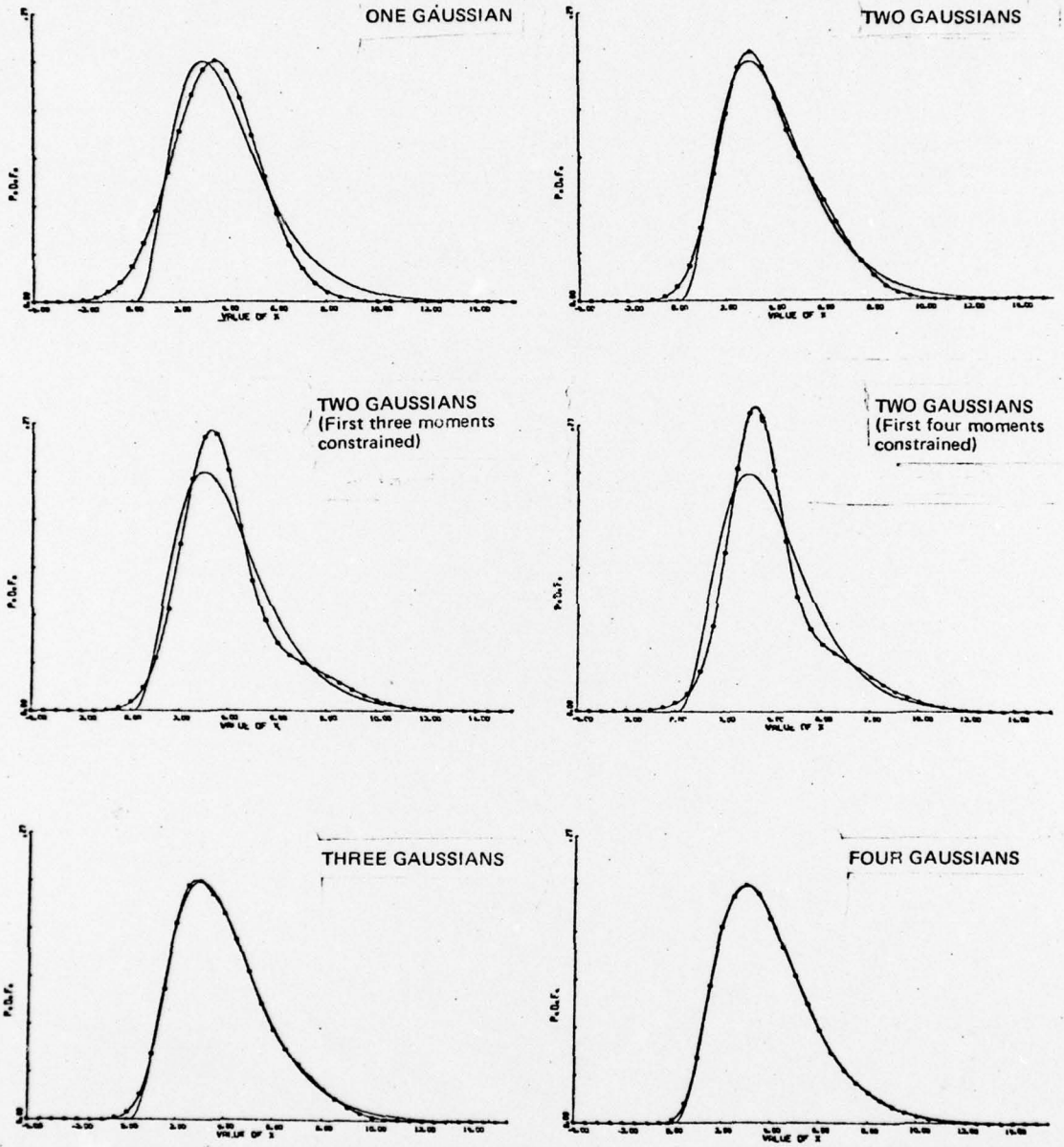
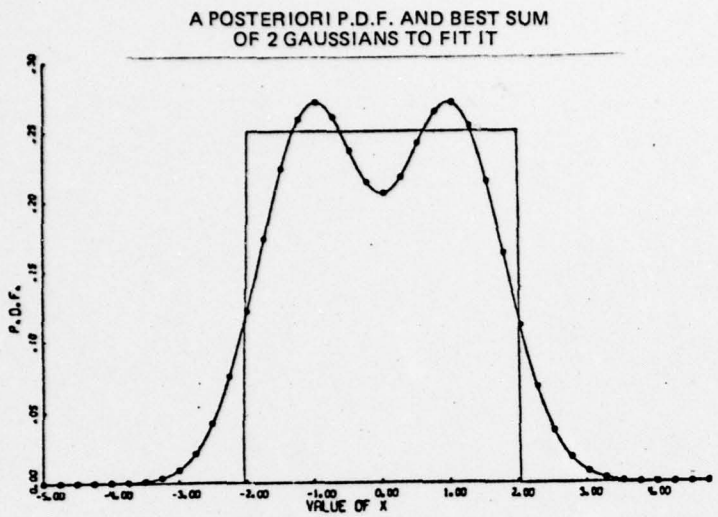
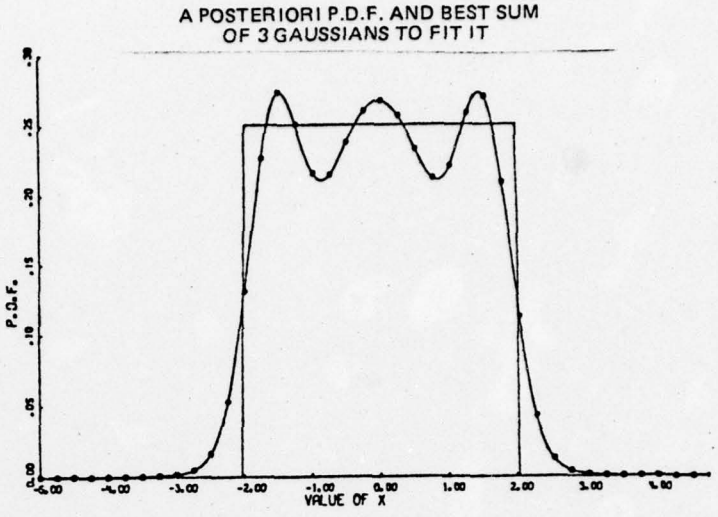
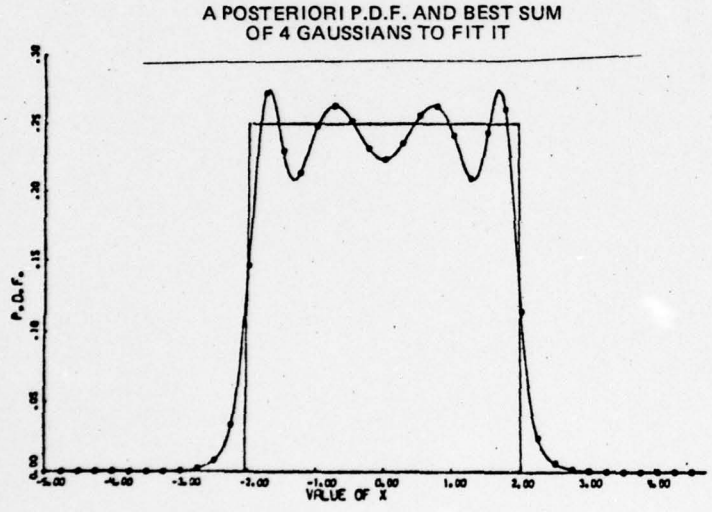


FIGURE 2


ONE GAUSSIAN
TWO GAUSSIANS
TWO GAUSSIANS
(First three moments
constrained)
TWO GAUSSIANS
(First four moments
constrained)
THREE GAUSSIANS
FOUR GAUSSIANS


A POSTERIORI P.D.F. AND BEST SUM
OF 2 GAUSSIANS TO FIT IT
P.D.F.
VALUE OF X


A POSTERIORI P.D.F. AND BEST SUM
OF 3 GAUSSIANS TO FIT IT
P.D.F.
VALUE OF X


A POSTERIORI P.D.F. AND BEST SUM
OF 4 GAUSSIANS TO FIT IT
P.D.F.
VALUE OF X

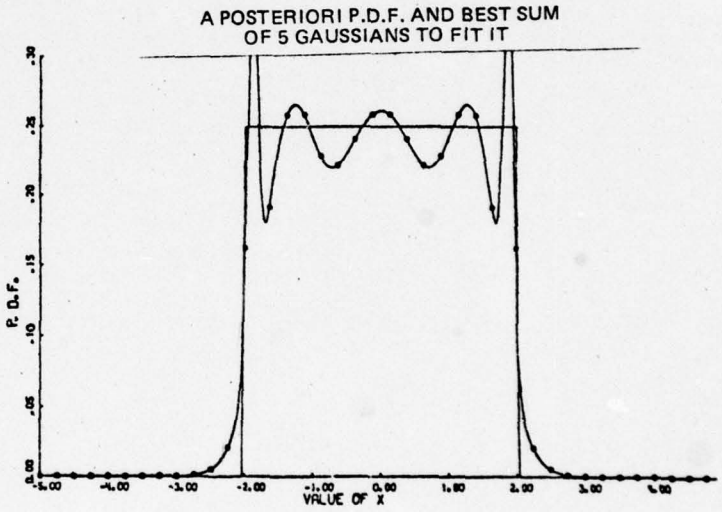

A POSTERIORI P.D.F. AND BEST SUM
OF 5 GAUSSIANS TO FIT IT
P.D.F.
VALUE OF X

FIGURE 3

FIGURE 4

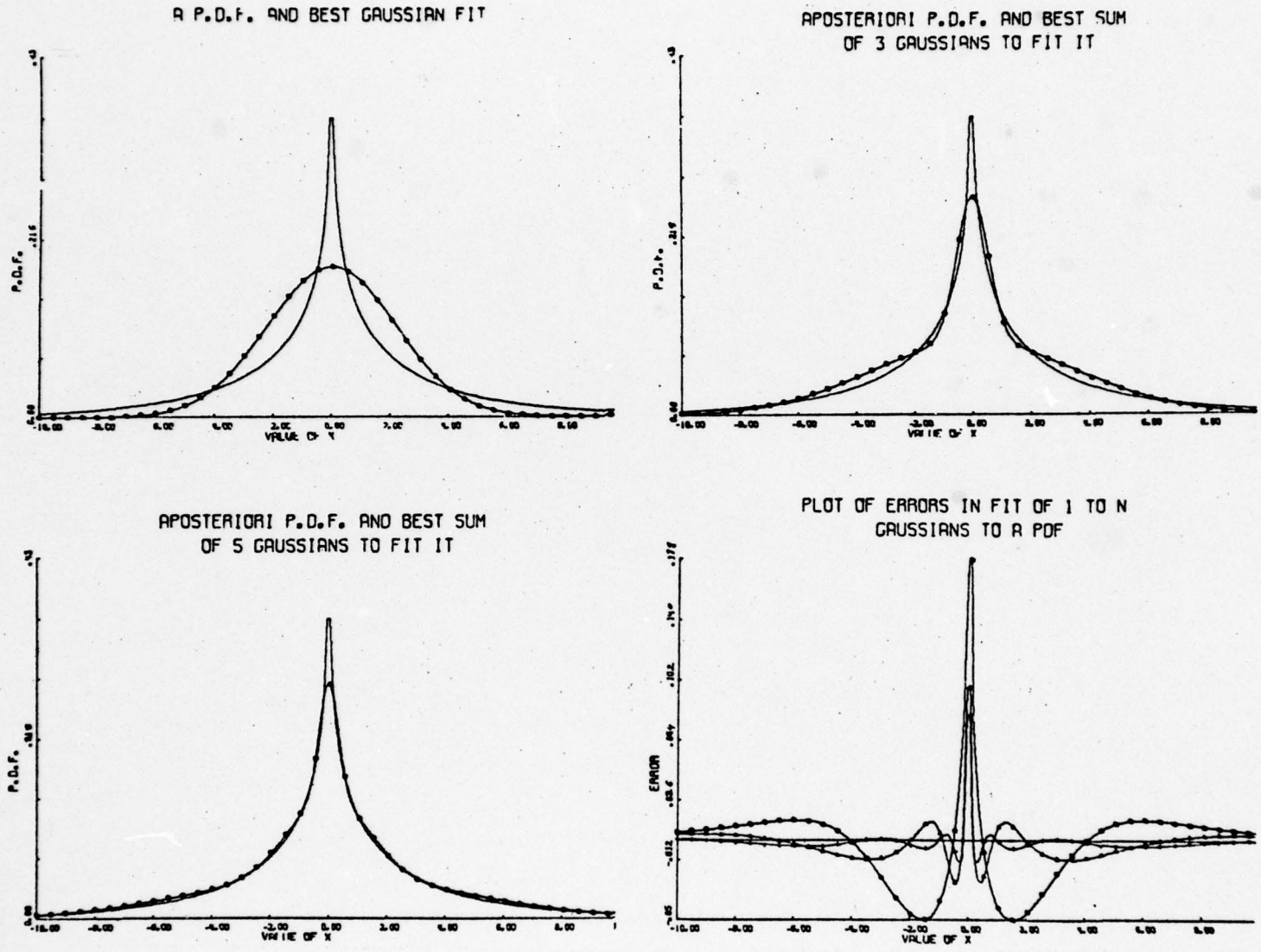
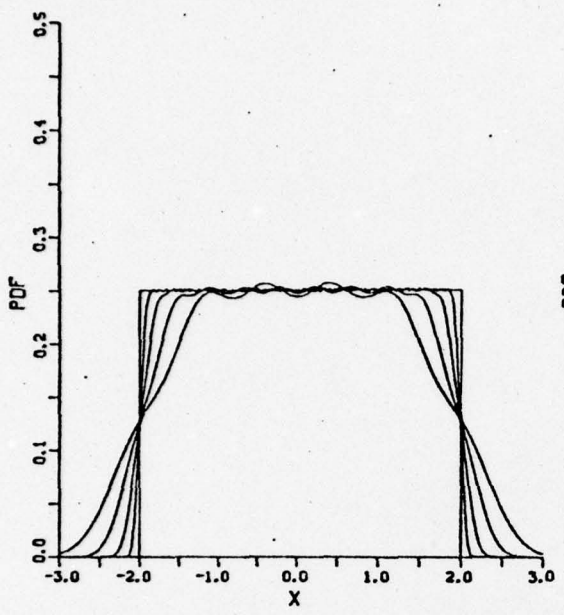
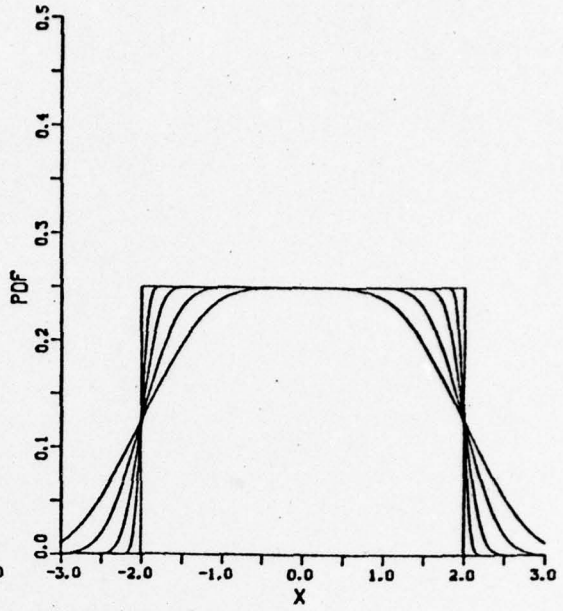

A P.D.F. AND BEST GAUSSIAN FIT
P.D.F.
VALUE OF X
APOSTERIORI P.D.F. AND BEST SUM
OF 3 GAUSSIANS TO FIT IT
P.D.F.
VALUE OF X
APOSTERIORI P.D.F. AND BEST SUM
OF 5 GAUSSIANS TO FIT IT
P.D.F.
VALUE OF X
PLOT OF ERRORS IN FIT OF 1 TO N
GAUSSIANS TO A PDF
ERROR
VALUE OF X

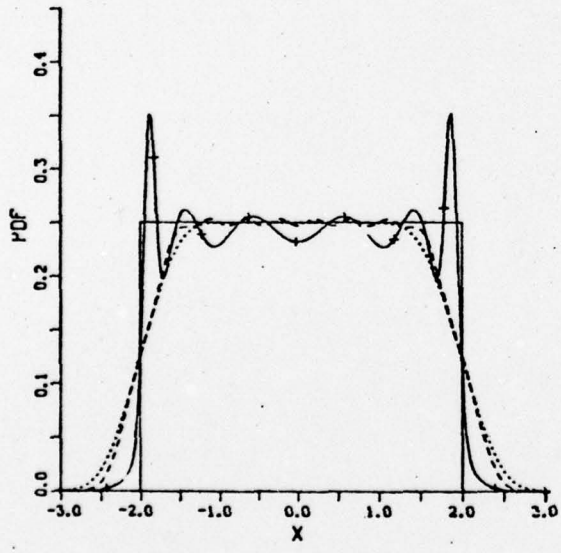
FIGURE 5



5a. Best theorem fit, 6, 10, 20 and 49 term approximations



5b. Smoothed theorem fit, 6, 10, 20 and 49 term approximations.



5c. L^2 search fit comparison Gaussians

+++ L^2 search fit

10-Term Theorem Fits

- - - - From Fig. 5a

. . . . From Fig. 5b

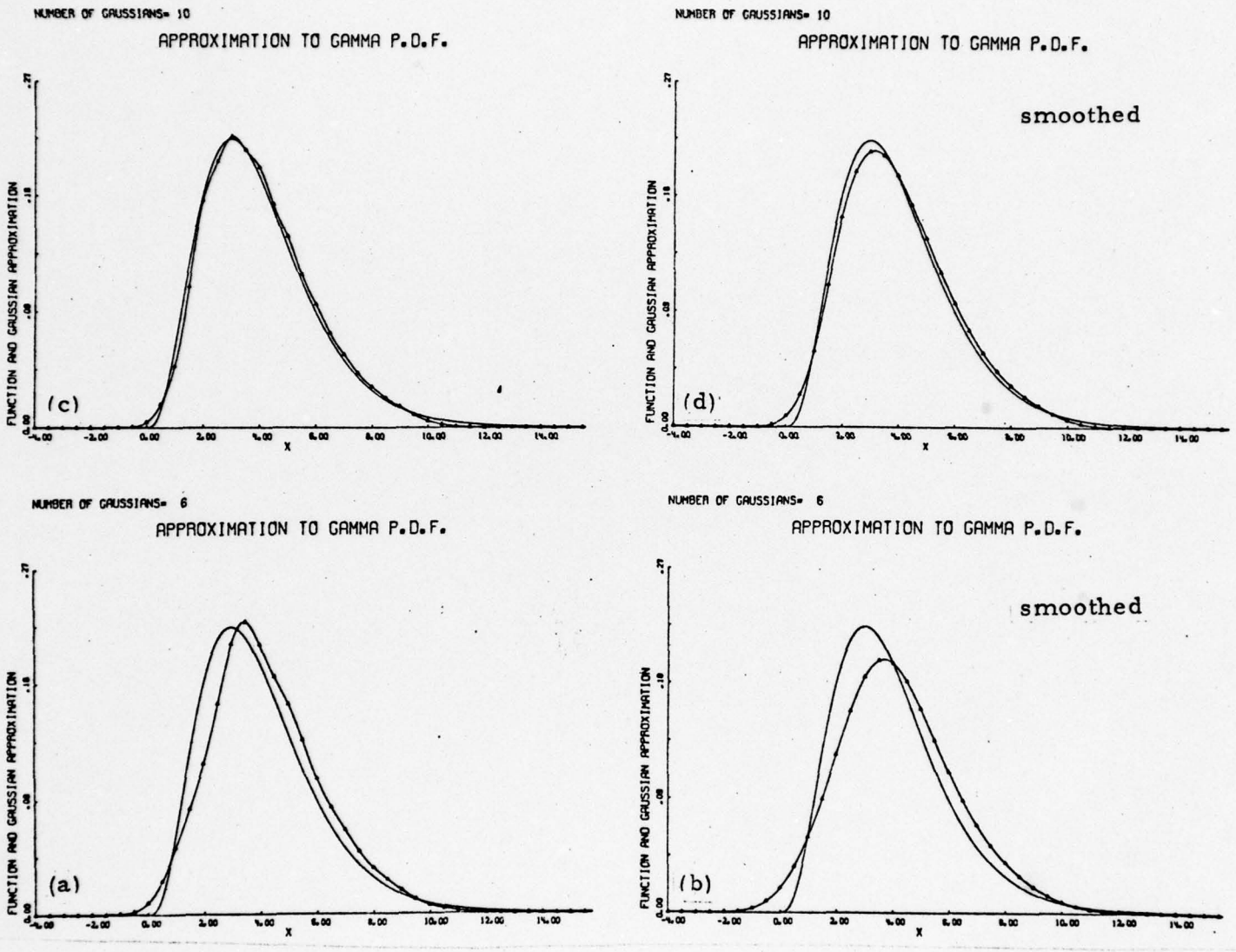
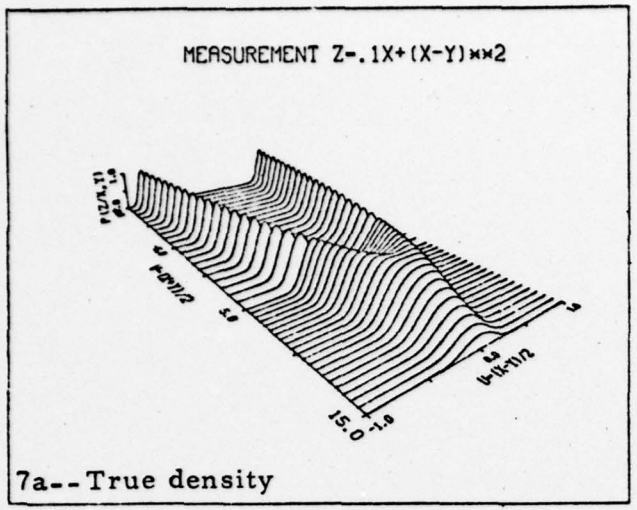
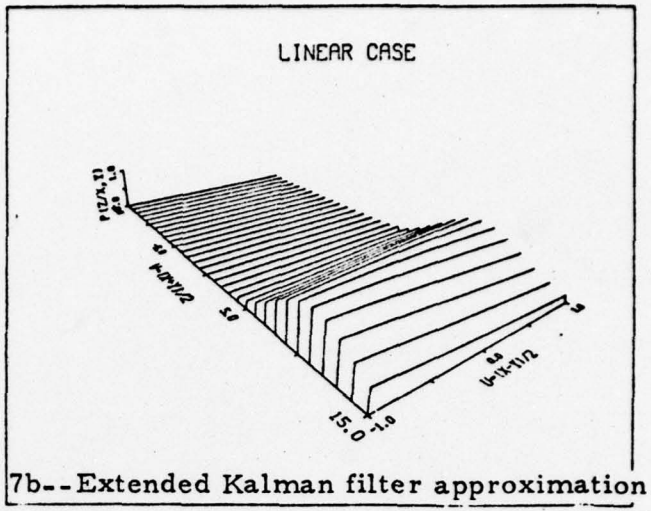

NUMBER OF GAUSSIANS= 10
APPROXIMATION TO GAMMA P.D.F.
FUNCTION AND GAUSSIAN APPROXIMATION
X
(c)
NUMBER OF GAUSSIANS= 10
APPROXIMATION TO GAMMA P.D.F.
smoothed
FUNCTION AND GAUSSIAN APPROXIMATION
X
(d)
NUMBER OF GAUSSIANS= 6
APPROXIMATION TO GAMMA P.D.F.
FUNCTION AND GAUSSIAN APPROXIMATION
X
(a)
NUMBER OF GAUSSIANS= 6
APPROXIMATION TO GAMMA P.D.F.
smoothed
FUNCTION AND GAUSSIAN APPROXIMATION
X
(b)

FIGURE 6

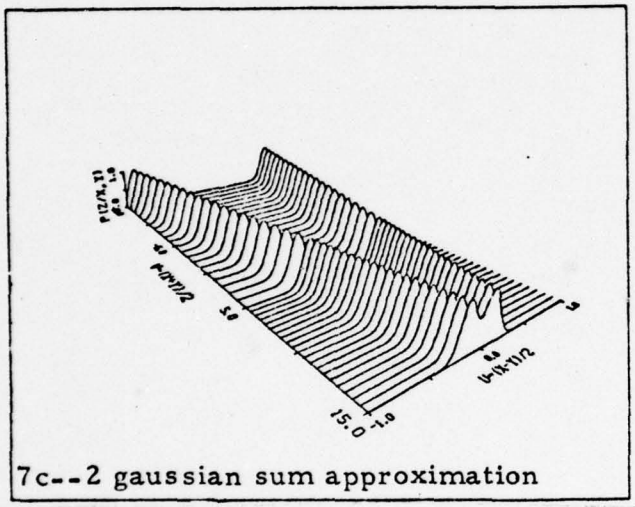
FIGURE 7



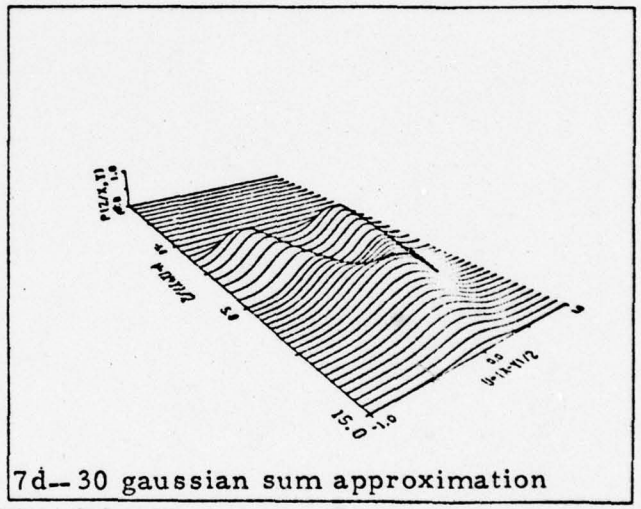
7a--True density



7b--Extended Kalman filter approximation

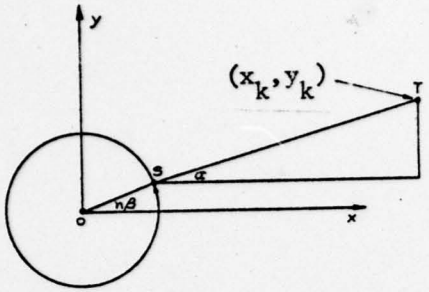


7c--2 gaussian sum approximation

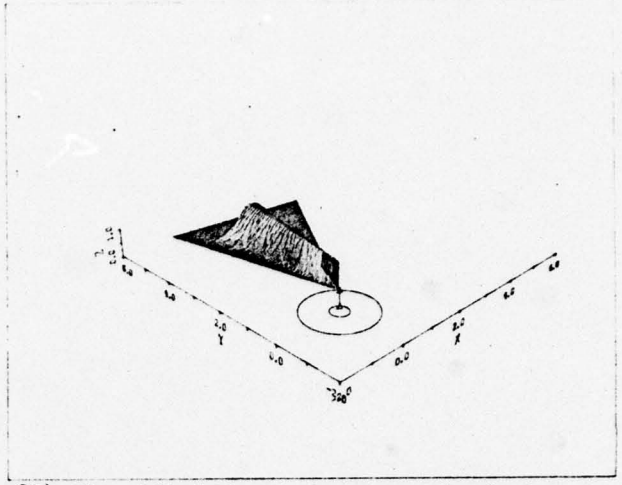


7d--30 gaussian sum approximation

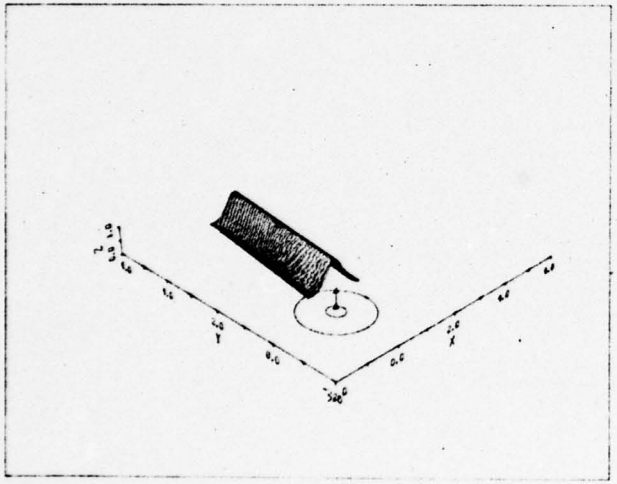
FIGURE 8



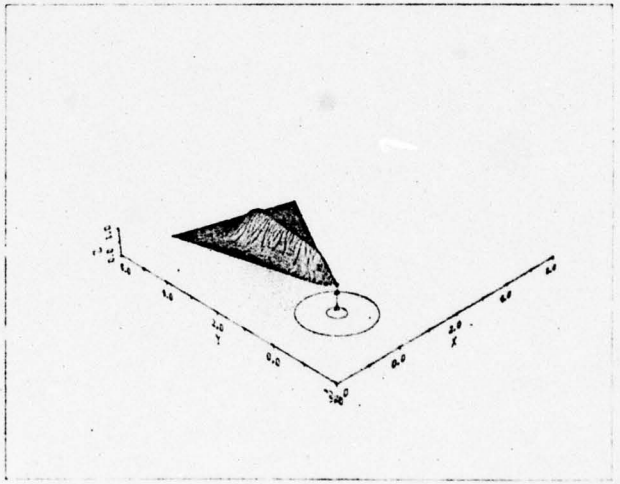
8a. Bearing-only tracking geometry.



8b. The likelihood density $p(z_k/\underline{x}_k)$.



8c. A Kalman filter approximation.



8d. A ten-term Gaussian sum approximation.

FIGURE 9

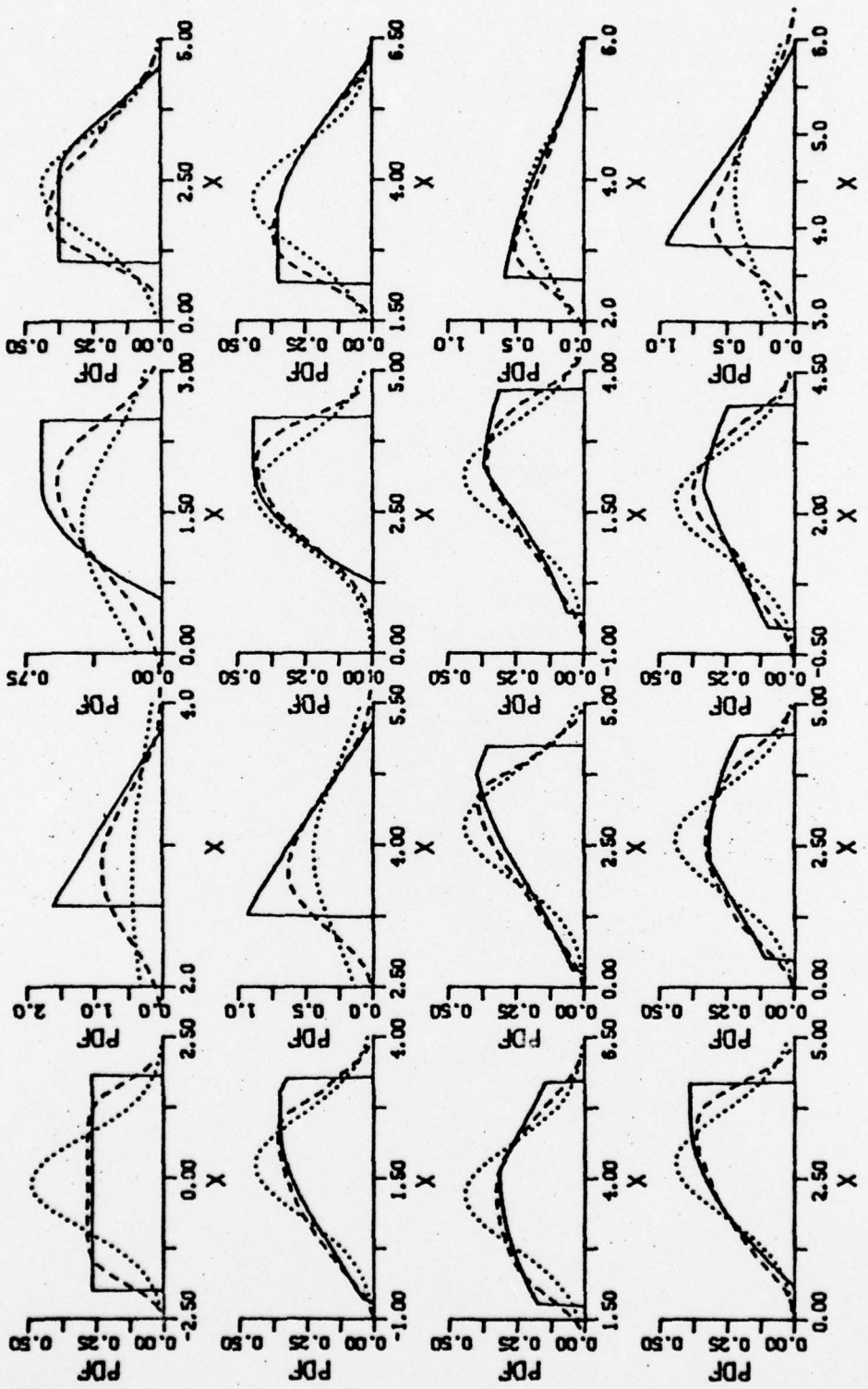


FIGURE 10

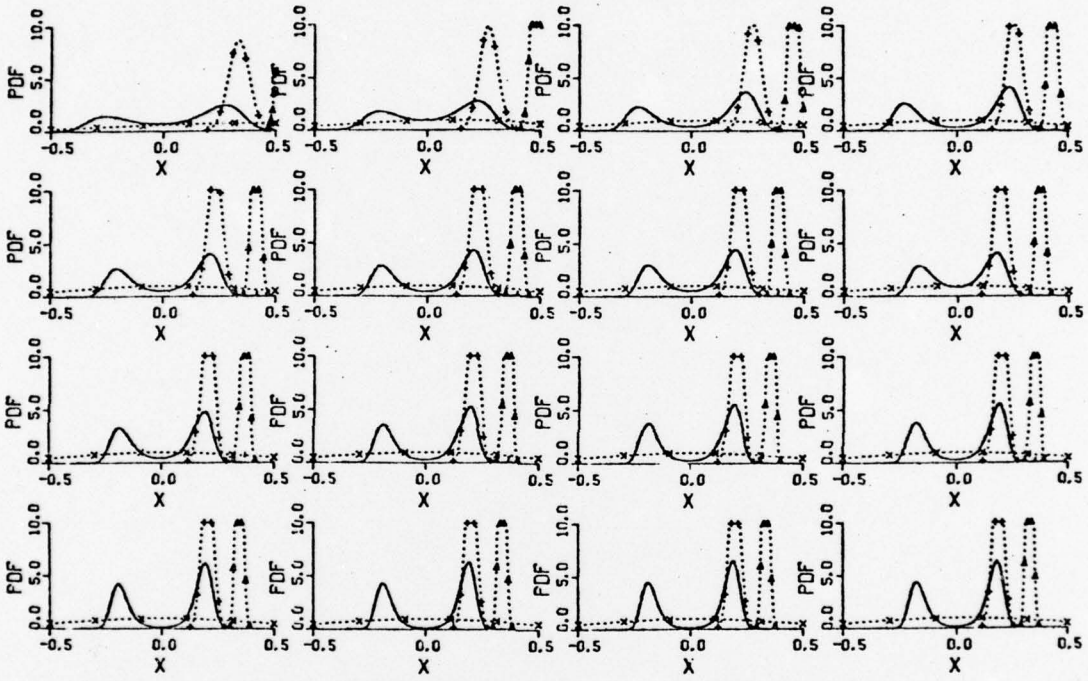


FIGURE 11

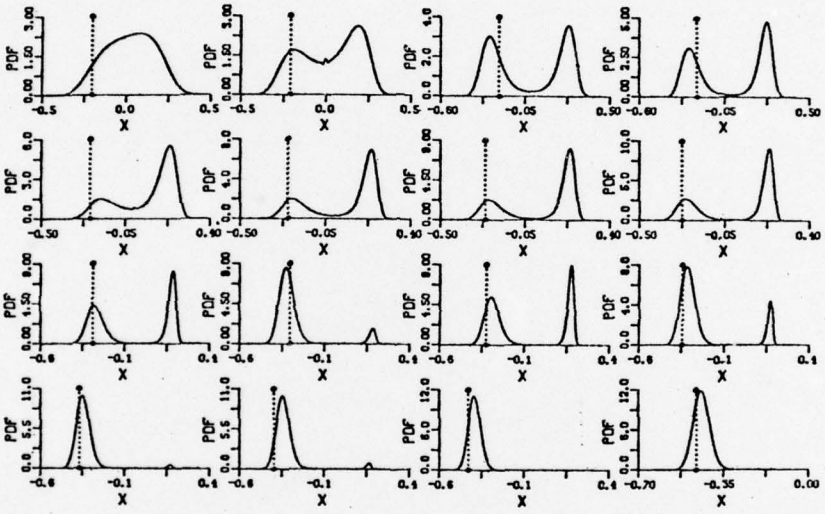


FIGURE 12

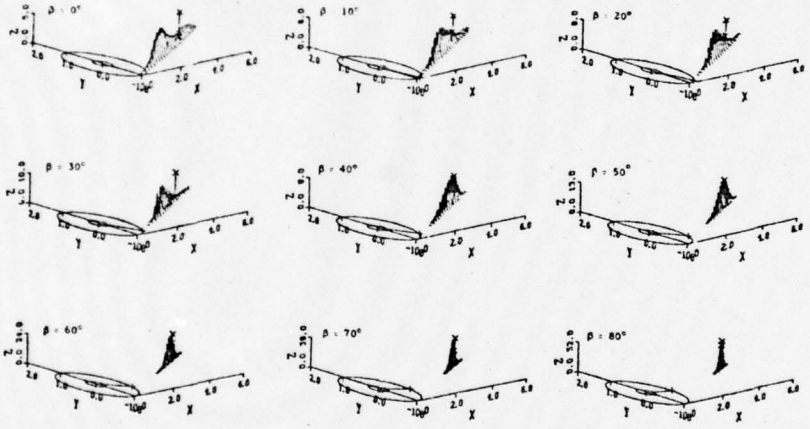
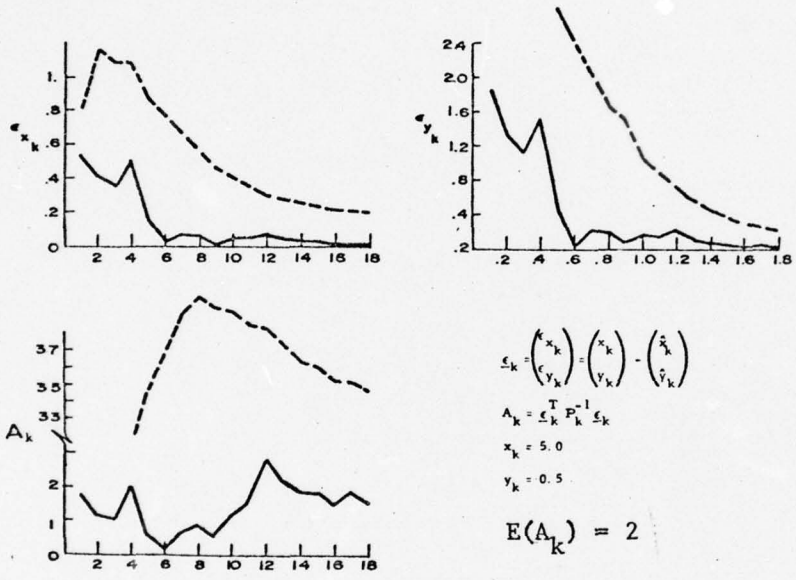


FIGURE 13



$$\underline{\epsilon}_k = \begin{pmatrix} \epsilon_{x_k} \\ \epsilon_{y_k} \end{pmatrix} = \begin{pmatrix} x_k \\ y_k \end{pmatrix} - \begin{pmatrix} \hat{x}_k \\ \hat{y}_k \end{pmatrix}$$

$$A_k = \underline{\epsilon}_k^T P_k^{-1} \underline{\epsilon}_k$$

$x_k = 5.0$

$y_k = 0.5$

$$E(A_k) = 2$$